

Query Match	Best Local Similarity	Score	DB 1	Length
Best Local Similarity 52.5%; Pred. No. 1.9e-11;				
Matches 623; Conservative 0; Mismatches 540; Indels 24; Gaps 3;				
QY 8	tcgatgcggtgacccggttcgatacccaacgagccgaagctgtacggtgtaaccctaccgct	67		
DB 963	TCGACGCGGCGCTCCGCGAGACCGCGCAACCGGTACGCGCGGATGCGCGAGAGAGCGGG	1042		
QY 68	ccctaccactggtctgtcgagcagacccgggtgcaacgttgcgccccacgggtcttgtagc	127		
DB 1043	TGCACCGCATCCGCAACGGGTGAACGGGCTTCGAAGCGCTGGCTGATCAACCGGTACAGAGACG	1102		
QY 128	tctcccgcttcggagacgttgcgagctgtctgcgcagacagcgcttcgccggacggagca	187		
DB 1103	TCMAACGAGCGGCTGCTGGACCCCAAGATTCGCCAAGACTTCGGGCGCAACGACAGATCA	1162		
QY 188	tccgcggttctcgagcagacactcgttcggggccgggtctgcgcgagatctgtcgagcagca	247		
DB 1163	TCGAACAAGAGCTGGCCGACCGCGAGCCAGGCGGGTTCTCGCCGCACTTCGCGCCG	1222		
QY 248	tcatctcgttctcaagagcaagacccgacacggcgcttgcgggggttggttcggccggcgt	307		
DB 1223	ACATGCTCAACACGACCCCGCGGACCAACCCGGTTGGCAAGCTGGTGGCAAGGCGGT	1282		
QY 308	tctcgcgcttcgagctgcgctgcgagacgggtgatactgcgcggcagccgcttgagcagactgc	367		
DB 1283	TCACCGCGCGGGGCTCGAGGGCTTCGCGCGCGGATCGAGCAGATCAACGCGACCTGC	1342		
QY 368	tgcggccggcccttcggccggggcgcgatgaggaagctgttcgaacagatctgcatacccgctgc	427		
DB 1343	TGCACCGCTTCGGCGGGCGGCTCCGAAGTGGATCTATGACGAGATTCCGCTTCGCGCTGC	1402		
QY 428	cgcctgcgagcgttctctcgagctgcgctgcgctgcgcggccgagactgggggagcgttcgggc	487		

Dd	1403	CCATACAGGTGATCTTCGGAGCTGAATGGGCGTGGAGACTCCTCGGCGGCAGCACTTCGGCT	1462
Oy	488	gctgttcgcgcagcttggaaacgaccttgaaaccggggccgaagcgcaggagaatcgcc	547
Dd	1463	CGT-----GGACCACAGTCTTGTTGAGCGGGTCTGCAGCCGGAGCGC	1504
Oy	548	gcgcgcacgcgcgcgatcgcgcgattccgcgcgaatacgtttggaacgagcccttcgcgagcgcc	607
Dd	1505	AGGCCACAGCGCTGGTGGCGAATGTCGAATCTGACCACTGATCCCAAGAAGCCCA	1564
Oy	608	gacgcgagcgcgcgaagaaccttcgcgcgatttgatgtctcgaacgcaccaagcacgcgcctca	667
Dd	1565	CCGACCCCGGCTGACGACCTGCTGTAACGGGCGTCTGGAAGCGGTGCAAGAGCGCACGCC	1624
Oy	668	tgaatcgaacaagagatcgtlcaagcagcgttgttaactcatcttcacgcgcacagagacgy	727
Dd	1625	TGAGCGAGGGGCGAGCTGATTCGCCATGAGTTCCTCTCTGTTCCCGGACAGGAGAGA	1684
Oy	728	tggcagcagcaggttggagaagccgcgtgcctgagcctctctgcgcgaaccgcgcagccgcgac	787
Dd	1685	CGGTCAACTGTATWCGGCAACTCTGCTGTCTGCTCTCTGGAACCCCGACACAGCTGCGCG	1744
Oy	788	tactccgcgcgcgcgcgcgaaccttcctgcgcacagccgcctgcagagatgltcgtcgttagaac	847
Dd	1745	CCTCGCGCAGCAACACCGCTGCTGCTGCTGCGCCGGCGCATGAGAGAACCTGGCGCTACGAGA	1804
Oy	848	cgtcgcgtgcagtccaacaacaccgcgcgcctcgaactgcagctgcagctgtgcgcgttcgcygc	907
Dd	1805	GCCCCGTGGCGCAACCGACGTTCCGGCACACCGCGAGCGGTCCGCTCCGCGACGACGTGG	1864
Oy	908	tgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc	964
Dd	1865	TGATCCCCGAGGAAGAGCTGCTGTGGGTGCGCTGCGGGCGGCGAACC GGAGGTGAGC	1924
Oy	965	ggtacgcacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc	1024
Dd	1925	GCTTCGAGGAGACCCCGACCGCTTCGACATACCCGGGAGACACACCGGACGCTGCGCTTCG	1984
Oy	1025	gcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc	1081
Dd	1985	GCCACGGGCACTTCATCTTCGTGGGCGCGCCCTTGACCGCGGTGGAGGCCCGCATGATCGCG	2044
Oy	1082	cgcgtgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc	1141
Dd	2045	TGGGCAAGGCTCTGAGAGCCCTTCGCGGACCTCGGATGGCCGCTGCGCGGACGACTGCG	2104
Oy	1142	ccatatcagcgcgcgcacacatgttcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc	1188
Dd	2105	GGTGGCGGTTCACAGCTCTGATGCGGGGCGCTGGAGAAGCTGCGCGGC	2151
RESULT	3		
AEO04675/c			
LOCUS	AEO04675	10060 bp	DNA linear BCT 30-AUG-2000
DEFINITION	Pseudomonas aeruginosa PA01, section 236 of 529 of the complete genome.		
ACCESSION	AEO04675.1		
VERSION	AEO04675.1		
KEYWORDS	GI:9948522		
SOURCE	Pseudomonas aeruginosa.		
ORGANISM	Pseudomonas aeruginosa.		
REFERENCE	Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae; Pseudomonas.		
AUTHORS	Stoyer,C.K., Phan,X.O., Erwin,A.L., Mizoguchi,S.D., Warrenner,P., Hickey,M.J., Brinkman,F.S., Huftnagle,W.O., Kowalik,D.J., Lagrou,M., Gardner,R.L., Goltzky,L., Tolentino,E., Westbrock-Wadman,S., Yuan,Y., Brody,L.L., Coulter,S.N., Folger,K.R., Kas,A.B., Lardig,K., Lim,R., Smith,K., Spencer,D., Wong,G.K., Wu,Z. and Paulsen,I.T.		
TITLE	Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pathogen		
JOURNAL	Nature 406 (679), 959-964 (2000)		
MEDLINE	20437337		

REFERENCE 2 (bases 1 to 10060)
AUTHORS Stover, C.K., Pham, X.-Q.T., Erwin, A.L., Mizoguchi, S.D., Warrenner, P.,
Hickey, M.J., Brinkman, F.S.L., Hutnagle, W.O., Kowalik, D.J.,
Lagrou, M., Garber, R.L., Goltzy, L., Tolentino, E.,
Westbrock-Wadman, S., Yuan, Y., Brody, L.L., Coulter, S.N.,
Folger, K.R., Kas, A., Larbig, K., Lim, R.M., Smith, K.A., Spencer, D.H.,
Wong, G.K.-S., Wu, Z., Paulsen, I.T., Reizer, J., Sailer, M.H.,
Hancock, R.E.W., Lory, S. and Olson, M.V.
TITLE Direct Submission
JOURNAL Submitted (16-MAY-2000) Department of Medicine and Genetics,
University of Washington Genome Center, University of Washington,
Box 352145, Seattle, WA 98195, USA
FEATURES
source 1. 10060
/organism="Pseudomonas aeruginosa"
/strain="PAO1"
/db_xref="taxon:287"
70. 1416
/gene="PA2472"
70. 1416
/gene="PA2472"
/codon_start=1
/transl_table=11
/product="probable MFS transporter"
/protein_id="AAG05860.1"
/db_xref="GI:9948523"
/translation="MPTPDKNNKIDVOAFIDAQVAFOLRVLSICFLIYAMGPDPA
AIGFIAPALAHMDSPALSPITLGAALAGLAGAPAGGLADRFGRKSVLLSVLFF
TUGSALGGLAAKVPALGWESVLLLGSGIPLASLPLMACLPESVRYFLARAPDAR
LRRLERLGRLLPDGMAGRELPADESDGTSPLROIIGAEIRGTLTLMATFEGLLII
YLTNLMLPTLIGTGSLGEANATISAMFOLGTLGALLGSAMDREDAHVSLSAVG
GALFIIIGLISLHSPALLALCVAGVGFCISQOVGANAADPEPTRSRATGVSMALG
LGRGISVGLSGGALLGLGFSGLLALLVYPALLAAVAVHRLGRRRPSFTTL"
1429. 2073
/gene="PA2473"
1429. 2073
/gene="PA2473"
/codon_start=1
/transl_table=11
/product="probable glutathione S-transferase"
/protein_id="AAG05861.1"
/db_xref="GI:9948524"
/translation="WDLYSEFNSSTYRVRIALALKGLDYVVVNLROGEOLRPADR
QRNPMGALPTLVADGRRESQSLAIDYDAVQPEPRLPLDPLHRAQLLEALIVAC
DIHPLNNVRPLKTLTGVLGIDADERQWYAHWYEGIAAETLLNHRGGAFFAGAA
GIVECCLVPOLANARWCGDIAPYPALLEGRCLALANQRASPERQDYPD"
2115. 3026
/gene="PA2474"
2115. 3026
/gene="PA2474"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAG05862.1"
/db_xref="GI:9948525"
/translation="MTTRIRASLIALPLLPFSFOATLPGMNLGQTSFLDKALPGT
LFOQVNIYEADPDANGHRRPLDNRLRLTLATSHFAIISOHLFGAYGAEIIVPV
LADPDEPHAGSANGNLIVSPFILQWPOTELLFRPYWQRLNALFOLPFGYDRH
AIDPNSNHSFNHYALTWEFAWGEASARLOYLLPGKRRDAPRLADXYLPGCAFH
ANYSVSLADAHMRILGSGYSLRQITDERVDGHDNRDSNERVDAIGPALPMGKSRLL
ELNAVHYEARNSSBSRLILRSQYF"
complement(3008. 4342)
/gene="PA2475"
complement(3008. 4342)
/gene="PA2475"
/codon_start=1
/transl_table=11
/product="probable cytochrome P450"
/protein_id="AAG05863.1"
/db_xref="GI:9948526"
/translation="MDDAFSEEGSAQPRHDAORPALAPRSDFGDIHTYHPDVFADPY
LRLIRSRAPVCDDQSIWMISRYADVSNACLDRRRSADPALGAGVYOGASAFNGH

QQLIPLARFYDNFMFLFNDRRTLRRLRFAPAFGPDVARNWEARIEVLVELLSILE
RRPDLIRDFAPDLITIRVAAELEFGPREDETCOLLWGRDLAAGIDLAAASHGAGCOINR
SAVAESDYLRORAGWSDSSPPSGAASIIIDGAMLEAGLIDIVAAVMVMAA
FETTIMVGNATLALTTHPDOLILRCGEISLANVVEILIRPDGAVRGVRCOTLEAVE
IGGORIPREKXWLSFLANRRPEMFARPDRIQDORANKQVAFARHPHICGLVYLA
RLEQCALGLVLRFPALASEPTDLKWRSSVFRLLERLPIVPESDAKTCE"
complement(4415. 5185)
/gene="dsbg"
/note="PA2476"
complement(4415. 5185)
/gene="dsbg"
/codon_start=1
/transl_table=11
/product="thiol:disulfide interchange protein dsbg"
/protein_id="AAG05864.1"
/db_xref="GI:9948527"
/translation="MRLPNLTITLIGLITLNLNAGLATAEELPAPITRWOKGARIILG
SPDAPDGLRGVAAEYONOGMALXYLPDCKHYTLTGHLFPAQCKDLSREPLERYVAPLA
KEMOKMEOSANLADGRADAPRVYTLFSDPNCPTCTMFMEQARWWDKQVLRHIVY
GIIREDSEAKSALIALASKDPPQKALHDHQAQASTLKEPLAIPAAVRKQLAGNMELME
SMGAATPAIFEYLNAGEWMQDQGAPOPDLAEIIGPR"
complement(5182. 6018)
/gene="PA2477"
complement(5182. 6018)
/gene="PA2477"
/codon_start=1
/transl_table=11
/product="probable thiol:disulfide interchange protein"
/protein_id="AAG05865.1"
/db_xref="GI:9948528"
/translation="MLSPSIGPIVYSLIOHLLFLTALGALLGWLAAAGGRNAPVYL
FLLIIGLIVARLAVVRYWYROYRQDFQOMLDIRDGGTLAPGLAAYLAFVAMR
PALRSIYGASLGLAFWLSLGLYIERGTLPBELSLRNAESVGLADFRPLV
NLNMSWCPCREMPVLOQAENPDVVFLEAOGESAETVRHFLQENLRNLNLS
DNGQLGGQVGSVALPTVYFYTAGRLLGSHLGELSRSLARYLEAREPAAAPATSS
SE"
complement(6018. 7781)
/gene="PA2478"
complement(6018. 7781)
/gene="PA2478"
/codon_start=1
/transl_table=11
/product="probable thiol:disulfide interchange protein"
/protein_id="AAG05866.1"
/db_xref="GI:9948529"
/translation="NRVILILMLLPELSAQPGEDLFAFRGATQGTDELPEVKAERF
TWERLDQOVOLRWQIAPGYVLYQKRLREFDGLDPAQLPQLEPPGSHSDEFGESQVVR
OSLELTLPAAAGOLRLGMOQCADAGLCYPPQSOALDLGGTGPAAAGTSEVAAEDGL
AGSLQAGNMLAMSLLFEGILGLAFAPCSILPMLDLGIVVSGAGPRGILLAGSYV
LSMALVYAGLGVVVALIGGNLOAMLOQPMILGSFPAALFEVIALPMFGFELQIPALR
DRLDGLSRKRGKGSILGAAALGSLIGVCPMNAPLAGALLYIAOTGNALHSGVLE
SLGIGIGMPLILLYVGSRRPLPKRPGPMNLYKGYRGLFICTANILRLPLGLGELWIG
LGGALLVYATRAALHTARGLARHVAVLEFAAGCTIGLWGAAMLGAAAGADDPWPRLDV
YAAANRGATPTASAEAPLVTSQPAEILDRLOAAKABQOWVLIDYVADWCVSCRIMK
QVFAPDGVLAALQGVRLRLDVTADNMAASRLRLHYQVPPPSLITWIGPEGEERRAR
LITGEVADGFLAHWOATRRRG"
7944. 8624
/gene="PA2479"
7944. 8624
/gene="PA2479"
/codon_start=1
/transl_table=11
/product="probable two-component response regulator"
/protein_id="AAG05867.1"
/db_xref="GI:9948530"
/translation="MHVLLITFDDDLINSGLVAGLNAAGLTVDVRYASADTQALLQYAR
FDVLVLIDGLPDEDELRLRLROQGVDPVLYVTARDAYTRVAGIQAGADYLLKRP
FDLRELTGARHLTDLRRSAGRCVNIIEHGRLSYDSTRETMWDGHPVLSRREOALLQA
LLNNRGRLTSGEQLKDSYYGFGDEVESNALNVHHLRRLKIGNALIVQTVAGLQYRLGP
ARGGGDDA"
8621. 9943
/gene="PA2480"
8621. 9943

```

/gene="PA2480"
/codon_start=1
/transl_table=11
/product="probable two-component sensor"
/protein_id="AAC05868.1"
/db_xref="GI:9948531"
/translation="MSLRRLRLTLTIGSAFTLLSLAAMMLCDLRCGMKRSFDRLLAA
SARVAVGILLOOLPOPQADQOTRTLSAEQIGIPGMACQVSLAGEVLAQSHATPEKPA
LDTASGFPDQETIGVMSRSTFVQAGLRVLTGRLDEREVRVLRVVAAPVTVL
LGSIALTLMICGAGGGLAPLKRTRDALTRRGADSLPELLEALPSLRKLVETQNLPR
IAQAIERRRLTGDAAHELRSPLVATITHLQVAMTGEAALAAIEGACDLHRT
LDOLLIAVRGSLSPEDGVQCSAEADAAVAMHDAGGDVMEELKADGISEAPLELV
PSYLAIALRLNDNALRHTAAGTRVLAETVAGCGARFVYRHDGPGTAEIADPLVLR
RFMHGKSGSGSLGALVQALVQRCQGRDPDSRADQLRVTLEPKRE"

```

Query Match	18.2%;	Score 220;	DB 1;	Length 10060;
Best Local Similarity	53.1%;	Pred. No. 1.1e-09;		
Matches 575;	Conservative 0;	Mismatches 480;	Indels 27;	Gaps 4;

[illegible][illegible]

FEATURES	Source
LOCUS	AB070947
DEFINITION	Streptomyces avermitilis polyketide-9 biosynthetic gene cluster.
ACCESSION	AB070947
VERSION	AB070947.1 GI:15824107
KEYWORDS	
SOURCE	Streptomyces avermitilis DNA.
ORGANISM	Streptomyces avermitilis
REFERENCE	Bacteria; Firmicutes; Actinobacteria; Actinobacteridae; Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
AUTHORS	1 (bases 1 to 24268) Omura,S., Ikeda,H., Ishikawa,J., Hanamoto,A., Takahashi,C., Shinose,M., Takahashi,Y., Horikawa,H., Nakazawa,H., Osonoe,T., Kikuchi,H., Shiba,T., Sakaki,Y. and Hattori,M.
TITLE	Genome sequence of an industrial microorganism Streptomyces avermitilis: Deducing the ability of producing secondary metabolites
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 98 (21), 12215-12220 (2001)
PUBMED	11572948
REFERENCE	2 (bases 1 to 24268)
AUTHORS	Ikeda,H.
TITLE	Direct Submission
JOURNAL	Submitted (01-SEP-2001) Haruo Ikeda, Kitasato University, Dept. of Microbial Chemistry, School of Pharmaceutical Sciences, 5-9-1 Shirokane, Minato-ku, Tokyo 108-8641, Japan
FEATURES	(E-mail:iked@mc.pharm.kitasato-u.ac.jp; Tel:+81-3-5791-6242, Fax:+81-3-3444-6197)
Source	Location/Qualifiers 1..24268 /organism="Streptomyces avermitilis" /db_xref="taxon:33903" /note="Actinobacteridae" 1..1875 /codon_start=1 /transl_table=11 /product="putative membrane protein" /protein_id="BAB65274.1" /db_xref="GI:15824108" /translation="MSAHFSPMIAVAVVLMFTVASYSRONVHAYPSGGDYEVAANTNLGPGVGLVASALIDVYLIVAVASIASIGISGLDPFVVEHAYSCAVAYITLTLMLNRRGKESGKFAIPTVEYVAGVGFIMAGLFRGIDLDTHNAPADYHIAEHQGLKALFYFLILIRAFSSGCAITGEVAISNCPAPRKSKNSLAAMLLATVMEGSGFLITVAALATVATVNRANENPDLINGVAVGDPYQNPISQVAVRGSGSFLITVAALAVLIVLAANTVINGEPPLIGSLIDADRITLPRLQRTGRDLASNGIVLLAGATLIVYCGADSTRITQIIVGVFSTLSQGVNRHNNRLATEKDDAKRRHVSRAINVFGAFETGTALVAVLTAKFTHGAWVALLGMCIFVYATMSAIRKHVRYVAEEIAAPGPSDS

[illegible]

/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="PUTATIVE SIGNAL PEPTIDE PROTEIN"
/protein_id="CAD17795.1"
/db_xref="GI:17431115"
/translation="MERQQLFAITAIVIAVAASVSLARDLNDERDPPGQALVHGH
ATAAAGR"
gene
complement(575..1135)
/gene="RSP0645"
/note="RS05577"
/note="RSP0577"
complement(575..1135)
/gene="RSP0645"
/function="miscellaneous; not classified regulator"
/note="Product confidence : probable
Gene name confidence : hypothetical
predicted by Codon_usage
predicted by Homology
predicted by Framep"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="PROBABLE TRANSCRIPTION REGULATOR PROTEIN"
/protein_id="CAD17796.1"
/db_xref="GI:17431116"
/translation="MKRSTPDPVDMKKVAIKERSPAAPQGAALSLDRTRAIL
KMLORDASISNVALADKVISPPACLRVERLKEIGCTRGVALLLEPKALDGTLMV
GVLDSTSPESFADPEKAVQKVCLECHVVGEPDYMVLVTRDNDYNRLHAEOL
YLRGVQIRTFKMLKQVLSITDLPFA"
1244..2260
/gene="RSP0646"
/note="RS05576"
1244..2260
/gene="RSP0646"
/EC_number="4.1.99.4"
/function="miscellaneous; hypothetical/global homology"
/note="Product confidence : probable
Gene name confidence : hypothetical
predicted by Codon_usage
predicted by Homology
predicted by Framep"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="PROBABLE 1-AMINOCYCLOPROPANE-1-CARBOXYLATE
DEAMINASE (ACC DEAMINASE) PROTEIN"
/protein_id="CAD17797.1"
/db_xref="GI:17431117"
/translation="MNLNKHPRHPIITFGPTPIQPLKRLSAHLGKVELYAKREDNSG
LAFGKNTRKLELYPEVLAVAGYDLVLSIGSIQSNOTQVAVAANHLGIKLVQENW
VNVADAVYDRVLENIILSRILGADVRLDAGFDIGIRPSWEQAMEDVRRAGKPPIPA
CGSEHPLGIGFVGAFAEYRQOEAERGFREDYIVCSVSTQACGMVGFADGGRAD
VGSIDSAKPEQTRQRIILRIARDTAKVLESGDITBEDVYDLTRGGPGPYGIPNGSGL
EARRLCARQEGMTIDPYVYEGKSMHGKIDVRKGEPEGSRLYLAHLGGPALNAVSL
FRNG"
complement(2336..3508)
/gene="RSP0647"
/note="RS05575"
complement(2336..3508)
/gene="RSP0647"
/EC_number="4.2.1.17"
/function="miscellaneous; hypothetical/global homology"
/note="Product confidence : putative
Gene name confidence : hypothetical
predicted by Codon_usage
predicted by Homology
predicted by Framep"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="PUTATIVE ENOYL-COENZYME A HYDRATASE PROTEIN"
/protein_id="CAD17798.1"
/db_xref="GI:17431118"

/translation="MSALPKPAEPAAAPEVLEFHVNRVALITLNRPSALNLSHG
VROLSAILAAGCDPRRIAAYVRGAGEGCPGAGGVVALYLAHQGAHGDGLOFP
VDYERLDYALHRRPKPVVALLDGLIMGGMDLGGARLRVTESSRIAMPETRLGFLP
DVGATHPLGAMPENALYIGLTGTLSGADLHGGLAVCVAPMALSFEEDRLRLSH
AGDPDMLQALRRVEPPCNIVPHAAIRARTPTWIVRYEDRSTIDRIYALRESLEED
PAREROMLQALITDAMATHSPMLVTEALRGSLTLACPFREIGIVARLEEED
FEGVRAHLVYKDRHPGMAPATLGIAGVERERVRHFLSSWRQAAHPHLHLGQPOAA"
gene
complement(3505..4305)
/gene="RSP0648"
/note="RS05574"
complement(3505..4305)
/gene="RSP0648"
/EC_number="4.2.1.17"
/function="small molecule metabolism; degradation; carbon
compounds"
/note="Product confidence : probable
Gene name confidence : hypothetical
predicted by Codon_usage
predicted by Homology
predicted by Framep"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="PROBABLE ENOYL-COA HYDRATASE PROTEIN"
/protein_id="CAD17799.1"
/db_xref="GI:17431119"
/translation="MIALDWLHDCTVALLTISRPANAFTPEGIRLOAQATVETLADP
RVRALVITIGDGPKEFSAGADINTFASGGRDVARBAARFGAEPALONARPVTAAN
GAMGGGLECALACDIRIAERHAOLAVETAVGLCCGCTQTLPMVLVGEEMARKML
TGERIDAQTLALRIGLIEVEVDTGAAEREAALAMARRVTLSPRAVTFKALHQARHG
PRTALAVERERFVDTLFGHPDREGVANFLKKRPPRHHTARPEALP"
complement(4302..5198)
/gene="mmsB"
/note="RSP0649; RS05573"
complement(4302..5198)
/gene="mmsB"
/EC_number="1.1.1.31"
/function="small molecule metabolism; amino acid
biosynthesis; isoleucine/valine"
/note="Product confidence : probable
Gene name confidence : putative
predicted by Codon_usage
predicted by Homology
predicted by Framep"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="PROBABLE 3-HYDROXYISOBUTYRATE DEHYDROGENASE
OXIDOREDUCTASE PROTEIN"
/protein_id="CAD17800.1"
/db_xref="GI:17431120"
/translation="MKITAFISLGMKAPMARNLKAGHALSVEDLDRAVQSLVDAGA
SAAPSPKAAAGACACAITMLPAAHVAHSVLTADGILAGIAKGVPIIDISTIDPSAK
LAALALAHNGPFDAPVSGGTGAAGATLFWYGGGAAYFEQVRPYLSMGNLIVIC
GEAGAGOGAKICNNLVILGITMAGVAEAMSGEALGIDPKVIGIINTSTRCNSDPY
NPFGVETIAPSSRGYTGFTDLMLDGLIAGAPAAASVQPYVLGALAQOIVQASS
CGAKIKDPSAVIRLYRPAGAR"
complement(5220..6767)
gene
Query Match 17.8%; Score 215.6; DB 1; Length 190050;
Best Local Similarity 52.2%; Pred. No. 7.1e-10;
Matches 529; Conservative 0; Mismatches 479; Indels 6; Gaps 2;
QY 185 gcatccgcgcgtctcggacgacctcgcggccgcggcgtcgcgagatcgtcgcg 244
|||||
Db 77566 GCATCCGCTGGATACGGCGACGAAGGCCCCAMCATCCCTGTTCCAGGGGTTCAACC 77625
|||||
QY 245 acatcatccgtctccaggacgacgacgacgacgacgacgacgacgacgacgacg 304
|| |||||
Db 77626 GGAATGTTCTGATGCTCATCCGCCATCGACACGCGCTGCGGCGCTGATGATGCAAG 77685
|||||
QY 305 cgttcgcgcgtcgcgcgtcgcgcgcgtgacgacgacgacgacgacgacgacgac 364
|||||

[illegible][illegible]

	complete cds; and unknown genes.
ACCESSION	AF127374
VERSION	AF127374.1
KEYWORDS	GI:4731328
SOURCE	
ORGANISM	Streptomyces lavendulae. Streptomyces lavendulae Bacteria: Firmicutes; Actinobacteriota: Actinobacteridae; Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces. 1 (bases 1 to 63734) Mao,Y., Varoglu,M. and Sherman,D.H. Molecular characterization and analysis of the biosynthetic gene cluster for the antitumor antibiotic mitomycin C from Streptomyces lavendulae NRRL 2564 Chem. Biol. 6 (4), 251-263 (1999)
JOURNAL	99201491
MEDLINE	2 (bases 1 to 63734)
REFERENCE	Mao,Y.Q., Varoglu,M. and Sherman,D.H. Direct Submission Submitted (10-FEB-1999) Microbiology, Uni. of Minnesota, Box196, 420 Delaware St. SE, Minneapolis, MN 55455, USA
AUTHORS	3 (bases 1 to 63734) Mao,Y.Q., Varoglu,M. and Sherman,D.H. Direct Submission Submitted (27-MAY-1999) Microbiology, Uni. of Minnesota, Box196, 420 Delaware St. SE, Minneapolis, MN 55455, USA
TITLE	Amino acid sequence updated by submitter
REMARK	Location/Qualifiers
FEATURES	1..63734
SOURCE	/organism="Streptomyces lavendulae" /strain="NRRL 2564" /db_xref="taxon:1914" 123..1367 /function="dehydrogenase" /note="ORF6" /codon_start=1 /transl_table=1 /product="unknown" /protein_id="AAD28446.1" /db_xref="GI:4731329" /translation="MRNTLFEIIGDIPPIGVPEQMVASVIRREYRGPHQAFNSEVVD VPKPGQALIVMAAGINYNWNASLGQPDVYSAROKSHDFHIGSGESGVSVA AVGGVGNVANGDEVILISGCOWETADIRLGADPMTGSQSVMGEBNYGSFPAPALA VDVOCHRPKPGLTWERACFTLLGATRAYROLCGMHDPNRPDPVLIMGAGGIGSM AIQTTRRAGGTTPAIVASDEERARI CRELAGOCTTNRLEDPMGLPLPIDGHEAMGRWT EGVGAFGREWEIVGESRSPTVLEHSGQAVTPTSMTLCDAGVVTCGGTTGNADI DLRLIMROKRLOCSHPANLRQCRRDIHMVANGDLPCLSWTGGFDIGKAHOLMDHN OHPOGNOAVLVNAPRTGLTFPA" 1428..1958 /note="ORF5" /codon_start=1 /transl_table=1 /product="linal homolog" /protein_id="AAD28447.1" /db_xref="GI:4731330" /translation="MDKIDITMSEREITRAVLORCRGLDRDELTVKSAYVEDAHNDHR GVINGNAHDEFKKQIVLPDRAYTGTLHLHSTIEIDGAGVSYCTATYTRRSSDCI KRVOGPRGYVRDREFRDGWVKIARRVLNDFSIAQEVPLDPAQAQGFNPSSHRDLDTTD ASYOVPLRGDPAATL" 2006..3229 /codon_start=1 /transl_table=1 /product="cytochrome P450 hydroxylase ORF4" /protein_id="AAD28448.1" /db_xref="GI:4731331" /translation="MSDPEQHAPLPRORCPSPPPELEELRTDPTISRMPADSFS GMULTRNADVAAALADBGVSSHCKGAKPWPNLAPEMARHEYLPGFLIMPDPHTFRY RRLTKRFEMTAIKRLERIEIQITYTEILDAMEAGGYDVLDQSPALPIPLYTELING DPPDAADLTDELGIAGIVLMIAGHETSANNLIGVTYLLFNADOMMLLRDISLDIR AVEELLRHOTIVOOGLPGVTRDMIEIGHOVKTGESLIASLPANRPBAVPPDDRLSDQ ITRRHNHFIAGHGHIICIGIMELARVEKRWQMRGLVTRPELTRMAAAPEDIRMRDQI VYGYNNPYTIDEAK"
CDS	3226..4458

gene
CDS
gene
CDS
CDS
CDS

```
/codon_start=1
/transl_table=11
/product="cytochrome P450 hydroxylase ORF3"
/protein_id="AAD28449.1"
/db_xref="gi:4731332"
/translation="MTGEPAVRGCGPFGAGAPAPVPFPHAPDRLEPPHPPWEPLEPERPL
ORVTLPGGEMAIARVODVRAVAPADREFSROLAVAGCAPREFLEPPORPDVAVLSGGP
DHARLRILYGVYFTPRVEDMRPLIQRADGLMEBMCPPADLVDFSLPFAVSMT
CELLVAPEDERKRCVWSDALLTTTAHTPAAPVARYMMQMDHYLGGLVAQRVRFAD
IGSLVATDEBDRKLEGEVLRLAEILTAGETASQIIPNLYVLFPHPLLIELRND
HDLIDPAVEBELREFYIGTVDFGFPRTAIEDVELGCVYRAQETVVPMSGMANRPELF
TDPPDLDAARRPNPLIGFGAGPRHHQLAGQALRAVELQITLTTLFRRYRIRLVAEESL
SMKEGLAMWGNIMVYW"
4552. 5658
/function="esterase"
/notes="ORF2"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAD28450.1"
/db_xref="gi:4731333"
/translation="MSTIDWEHSTKEAGMDPAALRLRTDVPVARGAAOLCVNRGCT
VLDLSCSSGDSLFLVYVATKPAALVAHALRGLIGDRPAAEYWPFAFGKGD
VTVRIYVLDHRRAGVPGVRIGIYATRRPAALGMEVRSVDLEQSRKMGCEVAAYHFSFEE
ILGELVQVQVGTGRSEDFWTSLSLPAFLGDLDMGLPESAMPVHVPAARAHSEMPNQ
MSNRGYQAVTIPVSGALSGTAQMARFQOMMEGSSIDGIRVLRPELVEARRKPSNG
GIDALSKRPVMSCEMIGGCGPPRGISLVNCTSPDSAFGHAGTTSVYVADPTRE
LYALSLNIOEPFGAGITRLREVSDLAIGCEAG"
5992. 6849
/function="transcriptional activator"
/notes="ORF1"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAD28451.1"
/db_xref="gi:4731334"
/translation="MLNLPKMERANHPSPQVGLCPLEVRSAGAGTGAAGVSGIRV
RTLLAALRLAIGQAMSTERLIKEVWADNPATIDKRAAVALLRLRYLGNEGMWILT
RPSGQVLTDPDHLDAVAETLVREGRALAEGRAPVAARLITFALOOMGEPRPADNM
AISTVSOCTETELTENRSEAVOAHIDRALELHNOELVAGELRSLNAPLHEPHIDQM
LALVSGQOAEMLAAVMQLRALAENLGDIPGRQLOQELHLRLIPADNGLTSGPAP
AEPILVRS"
complement(6854. 7666)
/gene="mlt"
complement(6854. 7666)
/gene="mlt"
/function="aminoguanate dehydrogenase"
/codon_start=1
/transl_table=11
/product="Mlt"
/protein_id="AAD28452.1"
/db_xref="gi:4731335"
/translation="MEFLPAPAGSGATRLXAVLGDVQVAKAPGLINPLLSGSDMA
VVVPHAVARLAEVYEGTIRKIGNLGILVYVPAKALGGLADLPALNLIIGANMA
RREDPGRKAYNPFQGLGCVQSLQAAGHTVRRRNVLYGAGSAGATATALLMADANV
SVHTDRQVLDLRLGSRKRPDGIKALGEPDCLAPAVNATPLGNRSSDPLEPDA
FVRPVAVYDVMMRPHETALLSAATAGRRVNHGIIHMLEQVPCYRAFFGMP"
7895. 8842
/gene="mlts"
7895. 8842
/gene="mlts"
/function="kinase"
/codon_start=1
/transl_table=11
/product="Mlts"
/protein_id="AAD28453.1"
/db_xref="gi:4731336"
/translation="MTGPDGAGGGGVYTFRRPAGDVYFADDLIAAVRRLGTHREPP
LAGVVAAPATCDAGYVYRTWPRSPNAGNLITAFQOLRPLGTFVACADGGDLAAAE
SRACRKLTVYGGTIGGGIYHGGRAWPBGHGSEVEGVVVDNRSGPDCGRACQ
VQAAVSGATLIRRAAERGRGTGFDELASGARLAPWAAEVAVESAAALATAYVIGICE
LAHPLELYVGGGFAAGYGVAAVAHYERLIRPGTDPVVRVRAVLTGCRSLIMQALL
AREAHGRNRPESPESRVSSDVSSVSGYDFGVDRAVGRSD"
```

[illegible]

[illegible][illegible]

Oy	144	cgtagcgacgagtgatctgagagacgagctgttgcgcg-----gacgcgatacgcgcggttc	199
Db	144	CGGCGCCCAAGGTCTCTAACCATCCGGGGGTCCGCGCGACACCCTGGGAGGCGCCCAACT	203
Oy	200	gagacgcacctcgt-----cgagcccgggcgtctgcgcgagatcgtctgcgcgacatcat	251
Db	204	CTACGGGAAAGCGTACCGGACGCCCGCGCGGGGAGATCGCGAGAGACTCAGCCACCACT	263
Oy	252	ctgtctccaaagagagcccgacaacgcgcgcgtctgucguggtgtgtccgcgcgtttctc	311
Db	264	GCTTCACCTTCGACCCCGCGACCATACCCGCTGTGGCTGTGTGGCGCGGTTCAC	323
Oy	312	gcccctgcgcctgcgcgcgcgttgaaacgcgtgtatctgcgcgcacgcgttgagacacctgcgc	371
Db	324	CCCGCGCAGGTGGACGGCTGCAACCGCATTTACAAAGGATTCACCGAGGATTTGCTGA	383
Oy	372	gcccgcctgcgcgcgcgcgcgcgttgaagcgtgtgttcgaacgactgtgcgtacccgcgtgcct	421
Db	384	CGCGATTGGCGCGCGCGGACAGCGCGACCTGTATGGCCCACTTCGAGATTCGCGTGACCAT	443
Oy	432	ggcgcgcgtctctgcgcgtctgcgtgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc	491
Db	444	CGCGGTATTTTGACACTGCTGGGCATTTCCGAGGCGCAGCGCCGACACGCC-----	495
Oy	492	gtcgcgcgaacgttggagacgcaccccttgaaacccgcgcgcgcgcgcgcgcgcgcgcgcgcgc	551
Db	496	--CGCCAGTCTGGGAGCGCCAGCGGAGCACTCTGTCCCGAGAGAGGCCCGACGCGCTCG	553
Oy	552	ccaaacgcgcgacatcgcgcgagtttcgcgcgcacgcgtgaagcgcgcgcgcgcgcgcgcgcgc	611
Db	554	CC-----GATGCCAAGGTGCACACTCTGCGCGCTGTGCTCGAGGCTCAGACCCCGCA	605
Oy	612	tgaagcgcgcgaagacactgtctgcgttgaatgcacgcacccaagacgcgcgcgcgcgcgcgcgc	671
Db	606	GCGCGCGCAGCGTGTACAGCGCGGCTGTGTGCAGGCGCGCGAGAGCGCCAGTGTGAG	665
Oy	672	tcgcgaacgagatcgtcaagcaggttgttaacgttatcttaacccgcgcgcgcgcgcgcgcgcgc	731
Db	666	CGAAGCGGAACCTGTCTCCATGGCCCACTGCTGTGATGATGACAGCGGCTTCGAGACCACT	725
Oy	732	cagccaggttgagcaacgcgcgtgtctgaactctgtgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc	791
Db	726	GAACTATATGGGACACCGCGCTGTATCACTGCGGTGCMAACCGGAGCAACTGGGTTGCT	785
Oy	792	cgccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc	851
Db	786	GCGGGCCAGCGGAACTCTGTGCCAACCCCATGTGAAGAACTGTGTCCGCACAGTCC	845
Oy	852	ggttcagttccaaacccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc	908
Db	846	GGTGCGGCTTCGATGTGGCTTACACGTGGAAAGACGTGGAACTGGAACGGGTTCACT	905
Oy	909	gcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc	968
Db	906	TCCCGCGGGAATACATCTGTGTTCCAACTGACCGCCCAACAGTATGCGCGAGCGCTT	965
Oy	969	gcagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc	1028
Db	966	CGAGATCTCCAGCCGCTCTCACTTCACCGCAACCGGATGCGCATCTCGCTGAGGCTT	1025
Oy	1029	ggaggtgcgcgtactgcgcgtctgcgttcaactgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc	1088
Db	1026	CGGCGTGCACACTGCTCGGCGCTCTGCTGCGCGCGGTGGAGGGGCGGATGCCATCA	1085
Oy	1089	gcgcctcgt---gcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc	1145
Db	1086	GCGCGTGTCCGCGGCTTCCCGCAACCTCCAGTTGGGGGTGCCCAAGCGGAGCTGCAGT	1145
Oy	1146	tcagcgcgcgcacatgtctcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc	1187
Db	1146	GCTGCCGATCACCTTCTCCGCGCGCTGTATCACGCGCGGT	1187

RESULT 11	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM
AE004755	14759 bp	DNA linear BCT 30-AUG-2000	AE004755	AE004091	genome.	Pseudomonas aeruginosa.	Pseudomonas aeruginosa.
AE004755.1	GI:9949456	PA01, section 316 of 529 of the complete	AE004755.1	GI:9949456		Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;	
REFERENCE	AUTHORS	Pseudomonas.					
JOURNAL	MEDLINE	1 (bases 1 to 14759)					
AUTHORS		2 (bases 1 to 14759)					
		Stover, C.K., Pham, X.-Q.T., Erwin, A.L., Mizoguchi, S.D., Warrenner, P.,					
		Hickey, M.J., Brinkman, F.S.L., Hufnagle, W.O., Kowalik, D.J., Lagrou, M.,					
		Garber, R.L., Goltzy, L., Tolentino, E., Westbrock-Wadman, S., Yuan, Y.,					
		Brody, K., Coulter, S.N., Folger, K.R., Kas, A., Larbig, K., Lim, R.,					
		Smith, K., Spencer, D., Wong, G.K., Wu, Z., and Paulsen, I.T.					
		Complete genome sequence of Pseudomonas aeruginosa PA01, an					
		opportunistic pathogen					
		Nature 406 (6799), 959-964 (2000)					
		2 (bases 1 to 14759)					
		Stover, C.K., Pham, X.-Q.T., Erwin, A.L., Mizoguchi, S.D., Warrenner, P.,					
		Hickey, M.J., Brinkman, F.S.L., Hufnagle, W.O., Kowalik, D.J.,					
		Lagrou, M., Garber, R.L., Goltzy, L., Tolentino, E.,					
		Westbrock-Wadman, S., Yuan, Y., Brody, K., Coulter, S.N.,					
		Folger, K.R., Kas, A., Larbig, K., Lim, R.M., Smith, K.A., Spencer, D.H.,					
		Wong, G.K., Wu, Z., Paulsen, I.T., Reizer, J., Sater, M.H.,					
		Hancock, R.E.W., Lory, S. and Olson, M.V.					
TITLE	JOURNAL	Direct Submission					
FEATURES	source	Submitted (16-MAY-2000) Department of Medicine and Genetics,					
		University of Washington Genome Center, University of Washington,					
		Box 352145, Seattle, WA 98195, USA					
		Location/Qualifiers					
		1. 14759					
		/organism="Pseudomonas aeruginosa"					
		/strain="PA01"					
		/db_xref="taxon:287"					
		complement(203..808)					
		/gene="PA3326"					
		complement(203..808)					
		/gene="PA3326"					
		/codon_start=1					
		/transl_table=1					
		/product="probable Clp-family ATP-dependent protease"					
		/protein_id="AAG06714.1"					
		/db_xref="GI:9949457"					
		/translation="MKTDKDKREGSGSHGATKLMEXALKVRKRVYTGVDKMAKD					
		VYQADKHILASIDDPYIFWVFNPSGGHSGSDMIDAIRITPKYIMVIGSGSVASAGL					
		IYAAADKRNKSLDPNTRFLHPQSGHSDGSPNSNIETIRRELYRKRERLDRIFFAETQO					
		TPKISADTERDPFMLNMEAVQYGLVKNIIYSERETLPQG"					
		1289..8347					
		/gene="PA3327"					
		1289..8347					
		/gene="PA3327"					
		/codon_start=1					
		/transl_table=1					

[illegible]

SOURCE
ORGANISM Streptomyces griseus DNA.

REFERENCE
AUTHORS Ueda, K.
TITLE Direct Submission
JOURNAL Submitted (03-OCT-1998) Kenji Ueda, The University of Tokyo,
MEDLINE Faculty of Agriculture, Department of Agricultural Chemistry; 1-1-1,
COMMENT Fax: 03-3812-0544)
On Oct 6, 1998 this sequence version replaced gi:666020.
D45916:Submitted (13-Feb-1995).
location/Qualifiers

FEATURES
source
1 . 2795
/organism="Streptomyces griseus"
/db_xref="taxon:1911"
330 . 335
341 . 1561
/gene="orf-2"
341 . 1561
/note="orf-1"
/note="similar to bacterial cytochrome p-450"
/codon_start=1
/transl_table=11
/protein_id="BAA33494.1"
/db_xref="GI:3702260"
/translation="MENTSYONKEVNRNCPDYAHELEFPDROLTLTEEPVSRIIRM
YGGGEMALVTREEDVRYTTPDRFSRSVAIVGRDPKMPPTPIQASITILMDPPASSH
LRLVIACTPFRPEVOMRGOTORYVDLDLMEFEESGPGCVARVAPPLITICENAI
DIPEADRPRLRHAMTMNVGAAGKONAYRAKALRNYBELTDARRSGEGLDIST
ATRKDEDELDDDELAVANVLITGDGTITDGNATITLRPLSLSKAEFORP
PRLEELTHIPFKGVGIPRIALDEVLSGLIKGDVVHSIYLANDSKFFORP
ELPDDEPTIIPMTFGMHNCLGAPLATMLEVAFSTLTIRPALDLVPEDVSMNTW
TSIMRWLPALPVTW"
RBS 1585 . 1590
gene 1596 . 2714
/gene="orf-1"
1596 . 2714
/note="orf-1"
/note="similar to chalcone synthase"
/codon_start=1
/transl_table=11
/protein_id="BAA33495.1"
/db_xref="GI:3702261"
/translation="MATLCRPAIVPBEVITMOOTIDLARETHAGHPORDLVLIQI
TGVRHLIOPLEKTLAHPGEVNQVEAEARTRVPEVVRALNAETPESEIDLIY
LYSCFEMPSPSLTAIIINMGFPRETROQLDAIOCGACAGAALNRAHDCEVYPSNS
VASCFCISLCYOPTDIGVSLNSGLFDGALSAAVRRGGGGMRLBENGSHLPDSD
EWMSIAVBDTEFHFLDKRVGTGEMMLAVLIDLVLDIAGVSNPMDFIYAGSPR
IDDLCHEIDLPREMERYSRATLTREGNIASVYFDLARLFDDGGAESAQOLINGPF
PGTAAYANGSAKRGLEADNDVGRIDELDTLAGVALSG"

BASE COUNT 435 a 1051 c 908 g 401 t

ORIGIN

Query Match 16.8%; Score 203.6; DB 1; Length 2795;
Best Local Similarity 52.1%; Pred. No. 3,46-08;
Matches 586; Conservative 0; Mismatches 514; Indels 24; Gaps 5.

DB 448 GCCGGTCCCGCATCCGATTCGCGTAGCAGAGGCGACAGGCCGTGGCTGCATCCCGCA 507

OY 138 CCGAGCGTCCGCGGCTCTCGCGCAGCAGCGCTTCCGCCGAGCCGCGCATCCCGCGATT 197

DB 508 CGAGAGCGTCCGCGAGCGTACACACCGACGCGGTTTCAGCCGACGCGCTCTCCGCGCG 567

QY	198	ctgagacgaacctgcgtccgagccgagcgtgtctccgagatcgtccgagacatcatccgtgt	257
Db	568	GGACTTCCCCCGATGACCCCGAGCCGATCGTGCAG-----GGGAGTCCATCAACT	621
QY	258	ccagagacgaagcccgagaccagccgagcgtgcgagggatgtgtcgtcccgagcgtctccgctc	317
Db	622	CATGAGACCCCGCCGAGCAGCCGGCTGGCTGGGGGCTGGGGCCAAAGACTTCAACCCCGG	681
QY	318	cgagcgtcgcgcgtctgaaacgggtgtgtcgcgacacgtgtgacgaacctgcgtgcgcgc	377
Db	682	TCGGCTGGAACAATGGCGCGGGGACCCAGCCGCTGGTGGAGACGGCTGTCGACGAGAT	741
QY	378	ccctgagccggagcgagatgtgacgtgtgtcagacagctgtgcataccggagcgtgcgtgcgc	437
Db	742	GGAGGAGGAGGGCTCAACCGGGGGAATTGCTGCCCGGGTCTTCGGGC---GGCTGCGCT	798
QY	438	gtctctgcgcctcgtctgcgcctgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc	497
Db	799	GATCACCATTGCGAGGCACTGCACATCCCGAGCGGACCGGCGCTTGCTCGGGGCCA	858
QY	498	cgagatgtgagcgaagcccttgagccgagcgcgccagcgccagagaaatgtccgcgcgcgcgc	557
Db	859	CGCCATGACCATATATAACAGTGGGGGCGCGGGCAAGCAAGCAACCGGTGGCGCAAGGC	918
QY	558	ggcgatgcgcgagttcgcgcgaactagttgtgagaggggcctctcgagagcgcgcggtgaggg	617
Db	919	GGAGCT-----GCGGGCTACTTTCAGAGAGTGAACCGCGACCGCGCGCTTCCCGGG	972
QY	618	cgagcgagagacgtcgtgcgtgtgattgtctgagcccaagacccgagcgtgatgtgcga	677
Db	973	CGAGGACCTCATAGCACCTTGCCACCGCCCGGACGCGACCAACATGCTGGAGCAGCA	1032
QY	678	cgagatcgttcagcagcgtgtgtcagcttatcttcaaccgagccagagacggtgtgcagca	737
Db	1033	CGAGTGGGCGCTCATGGGCGATGCTCTCTCTATCAACCGGCCAAGACACACGACTTACCA	1092
QY	738	gtgtggcgaaagccggtcgtgacgtcgtgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc	797
Db	1093	GCTGCGCAACATCGCTACACCTCTCAACCGCGCGGACCTGCTGGGCTTCCCTCGGGG	1152
QY	798	ccgagccgagacgtgtgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc	851
Db	1153	CGAAGCGAGCGGGTCCCGGACACCTTGGAGAGCTGCTGCGCAATCCCTTCCGCA	1212
QY	852	gttgcagtcacaacacccgagcagctcgaagtcgaagtcgaagtcgtgcgcgcgcgcgcgcgc	911
Db	1213	GGCGTGGGCAATCCCGGATCGCCCTGAGAGAGTGGAGCTGTCGGGCTCTCATCA	1272
QY	912	ccgagccgagacgtcgtgtgtgtgtcgtgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc	971
Db	1273	GGCGGCGAGCGTGTGCACAGTGTCTTACTTACGCGGCAACCGGGACACCCCAAGTTTCA	1332
QY	972	ccgagcccgagatttcgaacatcgagcgaagtcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc	1031
Db	1333	CGGTCCCGAAGAGTGAACCCCGAACCGGCGGACCATCCCAATATAGCTTGGGCTGGGG	1392
QY	1032	gtagcgctactcgtcctcgcgcgtctctactcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc	1091
Db	1393	CGCCCAACCACTGCGGGCGCGCGCGCGGCAACCATGAGACTGGAAGTGGCTTCTCCAC	1452
QY	1092	---ccgtgagccgaagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc	1148
Db	1453	GGTGTGACCCGCTTCCGGGCTCTGCGCTGTCTGTGTGAGTGTCCCGCGAAGAGCTTCTGTGGAA	1512
QY	1149	ggcgcgcacatgttccgcgggcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc	1192
Db	1513	CACGAGTGCATCTGGCTTACCGCGCTGCGCTGCCGTCAT	1556
RESULT 13			
E59713			
2795 bp DNA linear PAT 31-JAN-2002			

[illegible]

[illegible]

[illegible][illegible]

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 21, 2002, 15:13:00 ; Search time 207.34 Seconds
(without alignments)
10011.332 Million cell updates/sec

Title: US-09-724-797-35

Perfect score: 1209
Sequence: 1 gtcgtgcgtacgtccgtgac.....cgttcacgcgcggcggttga 1209

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

N_Geneseq_032802:*
1: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:*
2: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:*
3: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:*
4: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:*
5: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:*
6: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:*
7: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:*
8: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:*
9: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:*
10: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:*
11: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:*
12: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:*
13: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:*
14: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:*
15: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:*
16: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:*
17: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:*
18: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:*
19: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:*
20: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:*
21: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:*
22: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:*
23: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:*
24: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	214.4	17.7	1233	21	AAC55790 Hydroxylase orf3
2	214.4	17.7	12249	21	AAC55840 Complete Mitomycin
3	214.4	17.7	18331	21	AAC55857 Complete nucleotid
4	212.2	17.6	6085	18	AAT70153 S.longisporoflavus
5	204.4	16.9	1257	22	AAAF81357 Quorum sensing con
6	203.6	16.8	2795	22	AAAF81296 Streptomycetes grise
7	203.6	16.8	2795	22	AAAF81954 Streptomycetes grise
8	195.6	16.2	50937	21	AAA09469 Streptococcus glab
9	194.8	16.1	21185	21	AAA63350 Streptomycetes globi

10	194.8	16.1	63164	21	AAA63348 Streptomycetes globi
11	194.6	16.1	1224	21	AAC55788 Hydroxylase orf4
12	193.8	16.0	1700	14	AAO45569 Sequence soyC and
13	187.6	15.5	1400	12	AAO14548 Eryf gene. Saccar
14	180.8	15.0	4403765	22	AAI996683 Mycobacterium tube
15	179	14.8	5970	21	AAAF7635 Nucleotide sequenc
16	179	14.8	5970	21	AAZ56003 Contig 002 from co
17	175.8	14.5	1251	21	AAZ87301 S. venezuelae macr
18	169.4	14.0	1998	12	AAO11127 Sequence encoding
19	160.8	13.3	1188	21	AAC55828 S. lavendulae MmcN
20	160.8	13.3	53500	21	AAC55842 Complete nucleotid
21	160.2	13.3	2168	15	AAO73674 Mycinamicin IV hyd
22	157.4	13.0	1879	12	AAO11126 Sequence encoding
23	155.4	12.8	47981	22	AAF30757 Micromonospora meg
24	155.2	12.8	53789	19	AAV21187 Amycolatopsis medi
25	152.4	12.6	2918	21	AAZ43922 S. tendae nikkomyc
26	152.4	12.6	2919	21	AAZ24486 Streptomycetes tende
27	150.4	12.4	5821	18	AAT58686 DNA encoding S. fr
28	148.8	12.3	14806	22	AAO03809 Streptomycetes galli
29	148.8	12.3	77536	21	AAA14651 Nucleotide sequenc
30	146.8	12.1	1197	22	AAAF24894 Pimaricin biosynth
31	146.6	12.1	123401	22	AAD17186 Streptomycetes nous
32	145.4	12.0	12905	21	AAZ99101 S. fradiiae tylosin
33	143.8	11.9	12905	21	AAZ99100 Streptomycetes prist
34	142.4	11.8	1194	17	AAT58555 Streptomycetes prist
35	142.4	11.8	4496	17	AAT58553 Streptomycetes clavu
36	137.4	11.4	1227	22	AAD14511 Pimaricin biosynth
37	137.4	11.4	15079	22	AAD14499 Streptomycetes clavu
38	136.4	11.3	1194	22	AAAF24895 Streptomycetes clavu
39	135.8	11.2	15079	16	AAO91580 S. clavuligerus cl
40	132.2	10.9	125401	22	AAD17186 Streptomycetes nous
41	126.6	10.5	8169	19	AAV26609 Actinomadura hibis
42	126	10.4	1233	15	AAO61452 Cytochrome P450 SC
43	121.6	10.1	12441	21	AAD87284 S. venezuelae deso
44	121.6	10.1	13613	21	AAZ87319 Nucleotide sequenc
45	121.6	10.1	38506	21	AAAF5633

ALIGNMENTS

RESULT 1	
ID AAC55790	standard: DNA; 1233 BP.
AC AAC5790;	
XX	
DT 19-JAN-2001	(first entry)
XX	
DE Hydroxylase orf3 involved in MC biosynthesis.	
XX	
KW Mitomycin; biosynthesis; mitosane ring system; antibiotic; anti-cancer;	
KW anti-inflammatory; immune-enhancer; immunosuppressant; asthma;	
KW chronic obstructive pulmonary disease; respiratory inflammation;	
KW fungicide; pesticide; ds.	
OS Streptomycetes lavendulae.	
XX	
PM WO200053737-A2.	
XX	
PD 14-SEP-2000.	
XX	
PF 10-MAR-2000; 2000WO-US06394.	
XX	
PR 12-MAR-1999; 99US-0266965.	
XX	
PA (MINU) UNIV MINNESOTA.	
PA (SHER) SHERMAN D H.	
PA (MAOY) MAO Y.	
PA (VARO) VAROGLU M.	
PA (HEMM) HE M.	
XX (SHEL) SHELTON P C.	

[illegible][illegible]

KW Streptomyces griseus; actinomycete; polyketide synthase; rppA; drug;
 KM 1,3,6,8-tetrahydroxynaphthalene; melanine; agricultural chemical;
 KM malonyl CoA; ds.
 OS Streptomyces griseus.
 XX
 XX
 XX Key Location/Qualifiers
 FT CDS 1596..2714
 FT /ftag= a
 FT /product= "Actinomycetes polyketide synthase"
 PN JP2000342269-A.
 PD
 PD 12-DEC-2000.
 XX
 XX 02-JUN-1999; 99JP-0155526.
 XX
 XX 02-JUN-1999; 99JP-0155526.
 XX
 XX (MITR) MITSUBISHI RAYON CO LTD.
 PA
 DR WPI; 2001-303158/32.
 DR P-PSDB; AAB74843.
 XX
 XX Preparation of 1,3,6,8-tetrahydroxynaphthalene for use as a synthetic
 PT material for melanine and as an intermediate for drugs and agricultural
 PT chemicals, comprises using a microbe and an enzyme
 XX
 XX Claim 2; Page 6-7; 10pp; Japanese.
 PS
 XX
 XX The present invention describes a method for preparing
 CC 1,3,6,8-tetrahydroxynaphthalene comprising using a microbe cell
 CC containing an rppA gene encoding actinomycetes polyketide synthase, to
 CC convert malonyl CoA as the substrate or a substance which can produce
 CC malonyl CoA in the microbe cell to 1,3,6,8-tetrahydroxynaphthalene and
 CC recovering it. 1,3,6,8-tetrahydroxynaphthalene is used as a synthetic
 CC material for melanine and as an intermediate for drugs and agricultural
 CC chemicals. The present sequence encodes the actinomycetes polyketide
 CC synthase protein isolated from Streptomyces griseus.
 XX
 XX Sequence 2795 BP; 435 A; 1051 C; 908 G; 401 T; 0 other;

[illegible]

Db	799	gatacaccatctgcagagcacttcagacatcccccagagcgagaccggccctgtctccggccca	858
Qy	498	cgacgctgcagacggaacccctctgnaaccgggacgcacagcgagagacatgcgcgcgcacgc	557
Db	859	cgccatgcacatctgatgaaacgtctggggcgcgcgcagacagagacgctgcgcgcacagcgc	918
Qy	558	ggcgatctgcgcgaattcttcgcgcgaatacgtctgagacggggcccttcggcgaagcgcgctgagagc	617
Db	919	ggagct-----ggcgcgctactcttcagagagacttgaccgcgcgaaccgcgcgcgcctcccgag	972
Qy	618	cgcgcgagagacatctctgcgtctgatctgacgcgcacacagacgcgcgcgcctgatctgatccaca	677
Db	973	cgagagacatcatcagacacccctctgcgcacccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc	1032
Qy	678	cgagatcttcagacagcgtctgtaacgtctatctatctaacgcgcgcgcgcgcgcgcgcgcgcgcgc	737
Db	1093	cgagcgtgcgcgtcatgtgcgaatgtctctgtctcatcaccgcgcgcgcgcgcgcgcgcgcgcgcgc	1092
Qy	738	ggtgcgcgaacgcgcgcgtctgaacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc	797
Db	1093	gctcgcgcgaacatgcgcctaacacccctctgtcacccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc	1152
Qy	798	ccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc	851
Db	1153	cgaaacgcgaacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc	1212
Qy	852	ggtgcgaatcccaaacaccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc	911
Db	1213	ggcggtctgcgcatacccccgtatccgcctctgagagacgtctgacgtctctccatcaaa	1272
Qy	912	ccgcgcgaacgctgcgtctgcctcctgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc	971
Db	1273	ggccgcgcgaacgctgcgtctgcctcctcctcctcctcctcctcctcctcctcctcctcctcctcct	1332
Qy	972	ccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc	1031
Db	1333	cgctgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc	1392
Qy	1032	gattgcgactgactgcctgcgtctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcct	1091
Db	1393	cgccacacacatcctgcgtgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc	1452
Qy	1092	---cctgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc	1148
Db	1453	gctcgtcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc	1512
Qy	1149	gcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc	1192
Db	1513	cacgagcttcacatctgcgttaccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc	1556
RESULT	8		
AAAO9469			
ID	AAAO9469	standard; DNA; 50937 BP.	
XX	AAAO9469:		
XX	29-AUG-2000	(first entry)	
XX	Streptococcus	oleandromycin gene cluster.	
XX	oleandromycin:	oleandrolide; polyketide synthase; oleA; oleAII; oleAIII	
KW	PKS; type I; 8,8a-deoxyoleandrolide synthase; modular; ketosynthase;		
KW	acyl-transferase; acyl carrier protein; inactivated; polyketide;		
KW	macrolactone; antibiotic; methylide; erythromycin; ss.		
XX	Streptococcus	antibioticus.	
XX	OS		
XX	Key	Location/Qualifiers	
FT	CDS	152..1426	
FT	FT	/+tag- a	
FT	FT	/label= oleI	
FT	CDS	complement (1528..2637)	

Query Match	Best Local Similarity	Matches	Conservative	Mismatches	Indels	Gaps
QY 98	16.28	Score 195.6	DB 21	Length 50937		
Db 46624	16.28	Score 195.6	DB 21	Length 50937		
QY 158	16.28	Score 195.6	DB 21	Length 50937		
Db 46684	16.28	Score 195.6	DB 21	Length 50937		
QY 218	16.28	Score 195.6	DB 21	Length 50937		
Db 46739	16.28	Score 195.6	DB 21	Length 50937		
QY 278	16.28	Score 195.6	DB 21	Length 50937		
Db 46795	16.28	Score 195.6	DB 21	Length 50937		
QY 338	16.28	Score 195.6	DB 21	Length 50937		
Db 46855	16.28	Score 195.6	DB 21	Length 50937		
QY 398	16.28	Score 195.6	DB 21	Length 50937		
Db 46915	16.28	Score 195.6	DB 21	Length 50937		
QY 455	16.28	Score 195.6	DB 21	Length 50937		
Db 46975	16.28	Score 195.6	DB 21	Length 50937		
QY 515	16.28	Score 195.6	DB 21	Length 50937		
Db 47032	16.28	Score 195.6	DB 21	Length 50937		
QY 575	16.28	Score 195.6	DB 21	Length 50937		
Db 47080	16.28	Score 195.6	DB 21	Length 50937		
QY 633	16.28	Score 195.6	DB 21	Length 50937		
Db 47140	16.28	Score 195.6	DB 21	Length 50937		
QY 692	16.28	Score 195.6	DB 21	Length 50937		
Db 47200	16.28	Score 195.6	DB 21	Length 50937		
QY 752	16.28	Score 195.6	DB 21	Length 50937		
Db 47260	16.28	Score 195.6	DB 21	Length 50937		

Db 5424 aaccgacctgctgctccagcagccagagcatagacgtgtctctccctccggtggtc 5483
QY 219 cgggtctgtctc-gccagatcgtctgagcatcatctctgttccagagcagccagcaag 277
Db 5484 cgcactctctcgttcgggggtttaccagctcatctgctctacagcaaccgcccagtaacg 5543
QY 278 gccgctcgcgcgggtgtgtctgcgccgagcttctgcgctctgcgctctgcgctctgcgaac 337
Db 5544 gccggtctgcgcagagtggttgagctctgcgtttcaccccgccaaagtgctgcggatgtacg 5603
QY 338 cgggtatccgcgcagccgttgagcagacctgtc-----gcggccgcctctggccc 385
Db 5604 ggggtatcagagacatgacacgcacccgctccctccagaccagatgcccgcgctgcctccgcg 5663
QY 386 ggggggcagatgagcgtgtgtcagcagcgtgcgttaccgcgctgcgctgcgcgctgcctctc 445
Db 5664 gctcccgctgcacctcctacagcaggttcgcgcgcgcgctgcctccgctgcggtgatacagc 5723
QY 446 gctctcgcgcctgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 505
Db 5724 agatgatctgcgtcttccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 5783
QY 506 gacggacccttgagaccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 565
Db 5784 ccgtgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 5843
QY 566 ccgagctgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 625
Db 5844 acgagatgagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 5903
QY 626 acctgtgcgcgttgatgctgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 679
Db 5904 acctgtgcacacctgtcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 5963
QY 680 agatcgttcagcagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 739
Db 5964 aactgtatgagcagcatatgctgtgtctgcacagcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 6023
QY 740 tgggcaaacgcgcgtctgagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 799
Db 6024 tcggtcacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 6083
QY 800 ggcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 859
Db 6084 aaccgcgccttcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 6143
QY 860 ccaacacccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 919
Db 6144 tcacacagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 6203
QY 920 acgtcgtgtgtctgtcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 979
Db 6204 ccaagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 6263
QY 980 acgattctcagacatcagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1033
Db 6264 gcccttcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 6323
QY 1034 tgccttactgctcgtgtgtcttactcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1093
Db 6324 tgaactctcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 6383
QY 1094 tggcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1153
Db 6384 ggaatcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1179
QY 1154 gcacacgttccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1179
Db 6444 ttctacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 6469

RESULT 10

AAA63348
ID AAA63348 standard; DNA: 63164 BP.
XX
AC AAA63348;
XX
DT 06-MAR-2001 (first entry)
XX
DE Streptomyces globisporus C-1027 gene cluster.
XX
KW Eneidine C-1027 biosynthesis gene cluster; apoprotein; chromophore;
KM cancer; ds.
XX
OS Streptomyces globisporus.
XX
FH Key
FH Location/Qualifiers
FT complement (8..658)
FT /tag= a
FT /product= "ORF -7 protein"
FT complement (930..1478)
FT /tag= b
FT /product= "ORF -6 protein"
FT complement (1649..2713)
FT /tag= c
FT /product= "ORF -5 protein"
FT complement (2850..3237)
FT /tag= d
FT /product= "ORF -4 protein"
FT complement (3442..4971)
FT /tag= e
FT /product= "ORF -3 protein"
FT 5982..7479
FT /tag= f
FT /product= "glycerol phosphate transporter"
FT complement (7573..9900)
FT /tag= g
FT /product= "ABC transport/Ovra-1-like protein"
FT complement (9982..11349)
FT /tag= h
FT /product= "Na+/H+ transporter"
FT complement (11351..12835)
FT /tag= i
FT /product= "hydroxylase/halogenase"
FT 13012..14079
FT /tag= j
FT /product= "dNMP-glucose synthase"
FT complement (14212..14643)
FT /tag= k
FT /product= "CagA"
FT complement (14690..15922)
FT /tag= l
FT /product= "aminotransferase"
FT complement (15919..1653)
FT /tag= m
FT /product= "N-methyl transferase"
FT complement (16553..17924)
FT /tag= n
FT /product= "C-methyl transferase"
FT complement (18031..19191)
FT /tag= o
FT /product= "epoxide hydrazase"
FT complement (19267..19929)
FT /tag= p
FT /product= "anthranilate synthase II"
FT complement (19926..21407)
FT /tag= q
FT /product= "anthranilate synthase I"
FT complement (21424..22878)
FT /tag= r
FT /product= "coenzyme F390 synthetase"
FT complement (22875..23546)
FT /tag= s
FT /product= "iron-sulphur flavoprotein"
FT complement (23566..24702)
FT CDS


```

Db 47463 cgaactctctgctcgggggttcacacagctccatcgtcttaagaacacccgcgatacg 47522
OY 278 gccggtctgcgggggtgtgtcggccggcgcttcctcgcgcgtccgcgtcgcgcgtgtgaac 337
Db 47523 gccggtctgcgggggtgtgtcggccggcgcttcctcgcgcgcgtccgcgtcgcgcgtgtgaac 47582
OY 338 gccggtctgcgggggtgtgtcggccggcgcttcctcgcgcgcgtccgcgtcgcgcgtgtgaac 385
Db 47583 gccggtctgcgggggtgtgtcggccggcgcttcctcgcgcgcgtccgcgtcgcgcgtgtgaac 47642
OY 386 gccggtctgcgggggtgtgtcggccggcgcttcctcgcgcgcgtccgcgtcgcgcgtgtgaac 445
Db 47643 gccggtctgcgggggtgtgtcggccggcgcttcctcgcgcgcgtccgcgtcgcgcgtgtgaac 47702
OY 446 gccggtctgcgggggtgtgtcggccggcgcttcctcgcgcgcgtccgcgtcgcgcgtgtgaac 505
Db 47703 gccggtctgcgggggtgtgtcggccggcgcttcctcgcgcgcgtccgcgtcgcgcgtgtgaac 47762
OY 506 gccggtctgcgggggtgtgtcggccggcgcttcctcgcgcgcgtccgcgtcgcgcgtgtgaac 565
Db 47763 gccggtctgcgggggtgtgtcggccggcgcttcctcgcgcgcgtccgcgtcgcgcgtgtgaac 47822
OY 566 gccggtctgcgggggtgtgtcggccggcgcttcctcgcgcgcgtccgcgtcgcgcgtgtgaac 625
Db 47823 gccggtctgcgggggtgtgtcggccggcgcttcctcgcgcgcgtccgcgtcgcgcgtgtgaac 47882
OY 626 gccggtctgcgggggtgtgtcggccggcgcttcctcgcgcgcgtccgcgtcgcgcgtgtgaac 679
Db 47883 gccggtctgcgggggtgtgtcggccggcgcttcctcgcgcgcgtccgcgtcgcgcgtgtgaac 47942
OY 680 gccggtctgcgggggtgtgtcggccggcgcttcctcgcgcgcgtccgcgtcgcgcgtgtgaac 739
Db 47943 gccggtctgcgggggtgtgtcggccggcgcttcctcgcgcgcgtccgcgtcgcgcgtgtgaac 48002
OY 740 gccggtctgcgggggtgtgtcggccggcgcttcctcgcgcgcgtccgcgtcgcgcgtgtgaac 799
Db 48003 gccggtctgcgggggtgtgtcggccggcgcttcctcgcgcgcgtccgcgtcgcgcgtgtgaac 48062
OY 800 gccggtctgcgggggtgtgtcggccggcgcttcctcgcgcgcgtccgcgtcgcgcgtgtgaac 859
Db 48063 gccggtctgcgggggtgtgtcggccggcgcttcctcgcgcgcgtccgcgtcgcgcgtgtgaac 48122
OY 860 gccggtctgcgggggtgtgtcggccggcgcttcctcgcgcgcgtccgcgtcgcgcgtgtgaac 919
Db 48123 gccggtctgcgggggtgtgtcggccggcgcttcctcgcgcgcgtccgcgtcgcgcgtgtgaac 48182
OY 920 gccggtctgcgggggtgtgtcggccggcgcttcctcgcgcgcgtccgcgtcgcgcgtgtgaac 979
Db 48183 gccggtctgcgggggtgtgtcggccggcgcttcctcgcgcgcgtccgcgtcgcgcgtgtgaac 48242
OY 980 gccggtctgcgggggtgtgtcggccggcgcttcctcgcgcgcgtccgcgtcgcgcgtgtgaac 1033
Db 48243 gccggtctgcgggggtgtgtcggccggcgcttcctcgcgcgcgtccgcgtcgcgcgtgtgaac 48302
OY 1034 gccggtctgcgggggtgtgtcggccggcgcttcctcgcgcgcgtccgcgtcgcgcgtgtgaac 1093
Db 48303 gccggtctgcgggggtgtgtcggccggcgcttcctcgcgcgcgtccgcgtcgcgcgtgtgaac 48362
OY 1094 gccggtctgcgggggtgtgtcggccggcgcttcctcgcgcgcgtccgcgtcgcgcgtgtgaac 1153
Db 48363 gccggtctgcgggggtgtgtcggccggcgcttcctcgcgcgcgtccgcgtcgcgcgtgtgaac 48422
OY 1154 gccggtctgcgggggtgtgtcggccggcgcttcctcgcgcgcgtccgcgtcgcgcgtgtgaac 1179
Db 48423 gccggtctgcgggggtgtgtcggccggcgcttcctcgcgcgcgtccgcgtcgcgcgtgtgaac 48448

```

RESULT 11

AAC55788

ID AAC55788 standard; DNA; 1224 BP.

XX AAC55788;

AC

```

XX 19-JAN-2001 (first entry)
DT Hydroxylase orf4 involved in MC biosynthesis.
XX
XX
DE Mitomycin; biosynthesis; mitomycin ring system; antibiotic; anti-cancer;
XX anti-inflammatory; immunosuppressant; asthma;
XX chronic obstructive pulmonary disease; respiratory inflammation;
XX fungicide; pesticide; ds.
XX
OS Streptomyces lavendulae.
XX
XX WO200053737-A2.
XX
XX 14-SEP-2000.
XX
XX 10-MAR-2000; 2000WO-US06394.
XX
XX 12-MAR-1999; 99US-0266965.
XX
XX (MIND) UNIV MINNESOTA.
XX (SHER) SHERMAN D H.
XX (MAOY) MAO Y.
XX (VARO) VAROGLU M.
XX (HEM) HE M.
XX (SHEL) SHELTON P C.
XX
XX Sherman DH, Mao Y, Varoglu M, He M, Sheldon PC;
XX
XX WPI: 2000-601980/57.
XX
XX Novel nucleic acid molecule comprising mitomycin biosynthetic gene
XX cluster useful for cloning mitomycin biosynthetic genes for elucidating
XX the molecular basis of mitomycin ring system biosynthesis
XX
XX Example 1; Page 254; 399pp; English.
XX
XX This invention relates to isolated and purified nucleic acid molecules
XX from the mitomycin biosynthetic gene cluster. Mitomycins are a group of
XX natural products that contain a variety of functional groups, including
XX amino benzoxazine and axiridine ring systems. The S. lavendulae
XX mitomycin biosynthetic gene cluster comprises 47 mitomycin genes
XX spanning 55kb of DNA. The invention includes an expression cassette
XX comprising a mitomycin biosynthetic gene operably linked to a promoter,
XX and host cells transformed with the cassette. The nucleotide, and protein
XX sequences and the transformed host cells of the invention result in
XX antiasthmatic, anti-inflammatory, cytostatic, immunomodulatory, and
XX antibiotic activities. The nucleotide sequences are used to elucidate the
XX molecular basis for the biosynthesis of the mitomycin ring system, as well
XX as to engineer the biosynthesis of novel natural products, e.g.
XX antibiotics, anti-inflammatory agents, anti-cancer agents,
XX immune-enhancers, immunosuppressants, agents to treat asthma, chronic
XX obstructive pulmonary disease as well as other disease involving
XX respiratory inflammation, or cholesterol-lowering agents or as crop
XX protection agents (e.g. fungicides or insecticides) as well as
XX biopolymers, e.g., in packaging or biomedical applications, or to engineer
XX PHA monomer syntheses. Sequences AAC55782-C55881, AAC55815-C55849 and
XX AAC32485-B32542 represent mitomycin biosynthetic gene cluster DNA
XX sequences and encoded proteins. Sequences AAC55812-C55814,
XX AAC55550-C55556 and AAC55862-C55869 represent PCR primers used in the
XX cloning of the mitomycin biosynthetic genes.
XX
XX Sequence 1224 BP; 196 A; 492 C; 379 G; 157 T; 0 other;

```

Query Match

Best Local Similarity 16.18; Score 194.6; DB 21; Length 1224;

Matches 571; Conservative 0; Mismatches 504; Indels 24; Gaps 5;

```

OY 106 gccggtctgcgggggtgtgtcggccggcgcttcctcgcgcgcgtccgcgtcgcgcgtgtgaac 165
Db 121 gccggtctgcgggggtgtgtcggccggcgcttcctcgcgcgcgtccgcgtcgcgcgtgtgaac 180

```

QY 166 gagcgcttcgcccgcgacccgacatc-----cgccgttccttgagccacctgctcgagcc 219
 DB 181 cccgcgctcagctcgcacccgcgacccgcgaccccgcccttgccgacacccctgcgcccccgag 240
 QY 220 gggctcgtcgcgagatcgtcgcgacatcctcctgcctccagagagccagccacccagcc 279
 DB 241 atgcgcgcgcgagactacatcctgcgcgcgtcctcctcctcctcctcctcctcctcctcctc 300
 QY 280 cgcctcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 339
 DB 301 cgcctcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 360
 QY 340 gtcgctcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 399
 DB 361 aggatcgcgacagatcgtcctcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 417
 QY 400 gtcgctcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 459
 DB 418 ---gtcgcacctcgtcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 474
 QY 460 cccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 519
 DB 475 atgggcacatcgcctcagcagagagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 534
 QY 520 cgggcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 579
 DB 535 ctggacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 588
 QY 580 taagtcgagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 639
 DB 589 atgataagctcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 648
 QY 640 atgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 699
 DB 649 gccacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 708
 QY 700 agcttactcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 759
 DB 709 ctgattctcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 768
 QY 760 ctgctcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 819
 DB 769 ctgctcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 828
 QY 820 gccctcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 876
 DB 829 gccctcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 888
 QY 877 gaagtcgagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 936
 DB 889 accgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 948
 QY 937 gccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 996
 DB 949 ctgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1008
 QY 997 cgggacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1056
 DB 1009 cgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1068
 QY 1057 ctgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1113
 DB 1069 ctgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1128
 QY 1114 cgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1173
 DB 1129 cgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1188
 QY 1174 gccagcctcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1192
 DB 1189 tacacactccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1207

RESULT 12
 ID AA045569 standard; DNA; 1700 BP.
 AC AA045569;
 DT 04-DEC-1993 (first entry)
 XX
 DE Sequence soyC and soyC genes encoding cytochrome P450soy and
 DE ferredoxin-soy.
 XX
 KW P450soy: soyC gene; ferredoxin-soy: soyB gene; oxidation;
 KW haem protein; ss.
 OS Streptomyces griseus ATCC 13273.
 XX
 FH Key Location/Qualifiers
 FT CDS 149..1384
 FT /tag= a
 FT /product= P450-soy protein
 XX
 PN W09312236-A.
 XX
 PD 24-JUN-1993.
 XX
 PP 16-DEC-1992; 92NO-US10885.
 XX
 PR 16-DEC-1991; 91US-0807001.
 XX
 PA (DUPO) DU PONT DE NEMOURS & CO E. I.
 PI Omer CA, Sarislan PS, Trower MK;
 XX
 DR WPI: 1993-214178/26.
 DR P-PSDB; AAR38309.
 XX
 PT Constitutive expression of P450 SOY (SOYC) and ferredoxin soy
 PT (soyB) in Streptomyces - used for oxidn. of organic chemicals
 XX
 PS Example; Figure 2; 45pp; English.
 XX
 CC Cytochrome P450soy was purified from S.griseus ATCC 13273. Two
 CC similar forms of P450soy were isolated. P450soy-delta, is derived
 CC from P450soy by in vitro proteolysis during isolation. One of the
 CC tryptic peptide fragments of cytochrome P450soy and of of the
 CC P450soy-delta protein were subjected to automated degradation. The
 CC NH2 terminal sequences are given in R38306 and R38307. A mixture of
 CC oligos that consist of possible DNA sequences that could encode the
 CC AAs FGWVQCL of the tryptic peptide was made. It consists of the
 CC sequences in Q43290-23. The oligo mixture was end-labeled and used
 CC to probe the EMRI4 library of S.griseus DNA. Hybridising plaques
 CC were isolated and a 4.8kb SacI DNA fragment was isolated from one
 CC clone that hybridised to the oligo probe mixture. As segment of the
 CC 4.8kb fragment was sequenced and found to contain an ORF. Within
 CC this ORF was a section that corresp. exactly to the AA sequence
 CC determined from the cytochrome P450soy tryptic peptide (see Q45569,
 CC R38309) The gene encoding the P450soy protein was called soyC. Five
 CC nucleotides downstream from the stop codon for soyC another ORF was
 CC identified. This ORF encodes an apparent ferredoxin-like protein.
 CC The gene was designated soyB and the protein ferredoxin-soy.
 XX
 SQ Sequence 1700 BP; 254 A; 636 C; 574 G; 236 T; 0 other;

Query Match 16.0%; Score 193.8; DB 14; Length 1700;
 Best Local Similarity 51.4%; Pred. No. 1.7e-18;
 Matches 562; Conservative 0; Mismatches 507; Indels 24; Gaps 4;

QY 106 ggcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 165
 DB 308 gacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 367
 QY 166 gagcgcttcgcccgcgacccgacatcctcgttcgacgcacacctcgtcgggcccgagctg 225

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 21, 2002, 14:27:40 ; Search time 49:07 Seconds
(without alignments)
6051.985 Million cell updates/sec

Title: US-09-724-797-35

Perfect score: 1209

Sequence: 1 gtgtcgtgcgatgcctgtac.....cgtcacgcgcggcggttga 1209

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents NA: *
1: /cgn2_6/prodata1/ina/5A_COMB.seq: *
2: /cgn2_6/prodata1/ina/5B_COMB.seq: *
3: /cgn2_6/prodata1/ina/5A_COMB.seq: *
4: /cgn2_6/prodata1/ina/5B_COMB.seq: *
5: /cgn2_6/prodata1/ina/PCUS_COMB.seq: *
6: /cgn2_6/prodata1/ina/backfillseq1.seq: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	212.2	17.6	6085	4 US-09-029-603-4	Sequence 4, Appl1
2	195.6	16.2	50937	4 US-09-428-517-1	Sequence 1, Appl1
3	193.8	16.0	1735	1 US-08-102-863-10	Sequence 10, Appl1
4	193.8	16.0	1735	5 PCT-US92-10885-10	Sequence 10, Appl1
5	180.8	15.0	4403765	4 US-09-103-840A-2	Sequence 2, Appl1
6	179	14.8	5970	3 US-09-320-878-21	Sequence 21, Appl1
7	177.4	14.7	1251	4 US-09-105-537-38	Sequence 38, Appl1
8	177	14.6	1212	6 5212296-17	Patent No. 5212296
9	169.4	14.0	1998	6 5212296-8	Patent No. 5212296
10	157.4	13.0	1879	6 5212296-5	Patent No. 5212296
11	157	13.0	1221	6 5212296-16	Patent No. 5212296
12	142.4	11.8	1194	4 US-08-765-907A-9	Sequence 9, Appl1
13	142.4	11.8	4496	4 US-08-765-907A-6	Sequence 6, Appl1
14	137.4	11.4	1227	4 US-09-385-028-23	Sequence 23, Appl1
15	137.4	11.4	15079	4 US-09-385-028-1	Sequence 1, Appl1
16	121.6	10.1	13613	3 US-09-105-537-3	Sequence 3, Appl1
17	121.6	10.1	38506	3 US-09-320-878-19	Sequence 19, Appl1
18	121.6	10.1	44377	2 US-08-804-227C-7	Sequence 7, Appl1
19	121.6	10.1	44377	2 US-08-804-198-1	Sequence 1, Appl1
20	121.4	10.0	1209	4 US-09-105-537-21	Sequence 21, Appl1
21	116.8	9.7	68750	3 US-09-335-409-1	Sequence 1, Appl1
22	116.8	9.7	68750	4 US-09-568-102-1	Sequence 1, Appl1
23	116.8	9.7	68750	4 US-09-567-969-1	Sequence 1, Appl1
24	116.8	9.7	68750	4 US-09-568-480-1	Sequence 1, Appl1
25	116.8	9.7	68750	4 US-09-568-486-1	Sequence 1, Appl1
26	116.8	9.7	68750	4 US-09-568-472-1	Sequence 1, Appl1
27	116.6	9.6	4403765	4 US-09-103-840A-2	Sequence 2, Appl1

28	113.6	9.4	71989	4 US-09-443-501A-2	Sequence 2, Appl1
29	107	8.9	43280	2 US-08-804-227C-1	Sequence 1, Appl1
30	102	8.4	3013	2 US-09-096-982-6	Sequence 6, Appl1
31	102	8.4	3013	2 US-08-653-650A-6	Sequence 6, Appl1
32	98.8	8.2	2081	2 US-09-096-982-7	Sequence 7, Appl1
33	98.8	8.2	2081	2 US-08-653-650A-7	Sequence 7, Appl1
34	98.8	8.2	3196	2 US-09-096-982-4	Sequence 4, Appl1
35	98.8	8.2	3196	2 US-08-653-650A-4	Sequence 4, Appl1
36	98.2	8.1	8438	1 US-07-945-283-1	Sequence 1, Appl1
37	96.6	8.0	1269	1 US-08-396-218-1	Sequence 1, Appl1
38	96.6	8.0	1269	1 US-08-760-116-1	Sequence 1, Appl1
39	94.6	7.8	11220	4 US-09-105-537-32	Sequence 32, Appl1
40	94.6	7.8	36778	4 US-09-105-537-5	Sequence 5, Appl1
41	92.2	7.6	4257	2 US-09-690-473-1	Sequence 1, Appl1
42	92.2	7.6	4257	4 US-09-259-821A-1	Sequence 1, Appl1
43	92.2	7.6	4257	4 US-08-843-659-1	Sequence 1, Appl1
44	90.4	7.5	13842	4 US-09-105-537-30	Sequence 30, Appl1
45	88.6	7.3	12001	1 US-08-458-568A-11	Sequence 11, Appl1

ALIGNMENTS

RESULT 1
US-09-029-603-4
; Sequence 4, Application US/09029603
; Patent No. 6210935
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Engel, Natalie
; APPLICANT: Bietenhader, Jurg
; APPLICANT: Roupet, Christine
; APPLICANT: Pospiech, Andreas
; TITLE OF INVENTION: Staturosporin Biosynthesis Gene Clusters
; FILE REFERENCE: 4-20555/A/PCT
; CURRENT APPLICATION NUMBER: US/09/029,603
; EARLIER FILING DATE: 1998-03-20
; EARLIER APPLICATION NUMBER: PCT/EP96/03643
; EARLIER FILING DATE: 1996-08-19
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 6085
; TYPE: DNA
; ORGANISM: Streptomyces longisporoflavus
; FEATURE:
; NAME/KEY: misc_RNA
; LOCATION: (378)..(1665)
; OTHER INFORMATION: ORF
; FEATURE:
; NAME/KEY: misc_RNA
; LOCATION: (1747)..(2553)
; OTHER INFORMATION: ORF
; FEATURE:
; NAME/KEY: misc_RNA
; LOCATION: (2593)..(4011)
; OTHER INFORMATION: ORF
; FEATURE:
; NAME/KEY: misc_RNA
; LOCATION: (4013)..(4999)
; OTHER INFORMATION: ORF
; FEATURE:
; NAME/KEY: misc_RNA
; LOCATION: (5071)..(6085)
; OTHER INFORMATION: ORF
; US-09-029-603-4

Query Match 17.6%; Score 212.2; DB 4; Length 6085;
Best Local Similarity 52.0%; Pred.No. 1.3e+26;
Matches 525; Conservative 0; Mismatches 478; Indels 6; Gaps 2;

OY 90 cgaccgcgtgcacgcgtgcgcacgcgcgtgtctgtacgtccgcgttcgcgcgtgcg 149

[illegible]

RESULT 7
US-09-105-537-38
; Sequence 38, Application US/09105537A
; Patent No. 6265202

```

: GENERAL INFORMATION:
: APPLICANT: Sherman, D. H.
: APPLICANT: Liu, H.
: APPLICANT: Xue, Y.
: APPLICANT: Zhao, L.
: TITLE OF INVENTION: DNA encoding methymycin and pikromycin
: FILE REFERENCE: 600.438US1
: CURRENT APPLICATION NUMBER: US/09/105,537A
: CURRENT FILING DATE: 1998-06-26
: NUMBER OF SEQ ID NOS: 43
: SOFTWARE: FastSeq for Windows Version 3.0
: SEQ ID NO 38
: LENGTH: 1251
: TYPE: DNA
: ORGANISM: Streptomyces venezuelae
US-09-105-537-38

```

[illegible]


```
QY 99 gcaacgctggcggccacgggtctgtgtaacgtctccgcttcgcggaagctgctgcggtgtct 158
   || || || || || || || || || || || || || || || || || || || || || || ||
Db 129 GCTGTGGAGAGGACGACGAGGTGTGGTGTGACCTGCACGCGCGGGGCGCGGCGTCT 188
   || || || || || || || || || || || || || || || || || || || || || || ||
QY 159 cggagcgaagcagcttcgc-----cgagcggcatcgcgcggttctgtgacccagctgt 212
   || || || || || || || || || || || || || || || || || || || || || || ||
Db 189 GGGCGGACCGCGCTTCACCGCGGTGACGAGCGCGCGGCTTCGCGATCTGACCGCGCAC 248
   || || || || || || || || || || || || || || || || || || || || || || ||
QY 213 cgggcccggggtgtcgcgagatcgtcggaacatcatcagctgtccaggaagagccga 272
   || || || || || || || || || || || || || || || || || || || || || || ||
Db 249 CTCCTCACTGTGTGGCGCAACCGGAGTGGCGTCTTCATCCGATGAGACCCGCA 308
   || || || || || || || || || || || || || || || || || || || || || || ||
QY 273 ccacggcgcggtgctgcgggggtgtgtcggccgg--cgcttcgcgctccgcgctgcgcg 331
   || || || || || || || || || || || || || || || || || || || || || || ||
Db 309 GCATCTCCGGCTCGCTCATGTGTACCCGGGACTTCCTGGCCCGCGCGCGGAGCGCT 368
   || || || || || || || || || || || || || || || || || || || || || || ||
QY 332 tggaaacggtgatcgccggcaac--gtgaaagacctgtgcggcccgccctggccggag 389
   || || || || || || || || || || || || || || || || || || || || || || ||
Db 369 GCGCCCGCGGCTCGGGAGCTGCTGACGAGATCTGGGCGGCTGTGTGAAGGGGAGCG 428
   || || || || || || || || || || || || || || || || || || || || || || ||
QY 390 cggagatggagcgtgtgtgaagagctgtgacccgctgagcgtgcgctgcgctgcgct 449
   || || || || || || || || || || || || || || || || || || || || || || ||
Db 429 GCGGTCGACCTGTGTGCGGAGTGAAGATCCCGGTGCTGCGGGTATCACTGCT 488
   || || || || || || || || || || || || || || || || || || || || || || ||
QY 450 gctggcctgtccgcgcgcgactgtggggggtgtgcggtgtgtcgcggaagctgtggag 509
   || || || || || || || || || || || || || || || || || || || || || || ||
Db 489 CTTGCGC-----GCGGTGACGAGCCCGGAGTTTATGAGAGACCG 530
   || || || || || || || || || || || || || || || || || || || || || || ||
QY 510 gacccttgagcggggcgcaagcgcaagagatgtgcgcgcgcaagcgcgatgcgca 569
   || || || || || || || || || || || || || || || || || || || || || || ||
Db 531 CACCGGGGTCTCTATCGACCGCGGTACACCCGGGAGCGAGTCCGCAAGGCCCGGACCA 590
   || || || || || || || || || || || || || || || || || || || || || || ||
QY 570 gttcgcggaatagttgagggggtccttcgcaagcgcgcgcggtgtgagggcggaagact 629
   || || || || || || || || || || || || || || || || || || || || || || ||
Db 591 ACTGAGGGGTATCTGGGAGGTGTGTGAGAGCGGATCGAAGACCCGGGACCGACT 650
   || || || || || || || || || || || || || || || || || || || || || || ||
QY 630 gctggcgttgatgtcgaagcgcacagacg--cgacctgtatgtgtcgaagagatcgt 686
   || || || || || || || || || || || || || || || || || || || || || || ||
Db 651 GATCAGCGGCTGCTCATGACAGAGTGGCGGCGCGGACATCTGCGGATGAGAGATGCT 710
   || || || || || || || || || || || || || || || || || || || || || || ||
QY 687 cagcaagcgtgtgaagttcatcttcacgcgcgcgaagagcggtgtgcaagcaggtgtgca 746
   || || || || || || || || || || || || || || || || || || || || || || ||
Db 711 CCGAGTGTGCGGCTGCTGTGTGTGCGGTGACGGGACACCAAGCAGCGAGGAGCT 770
   || || || || || || || || || || || || || || || || || || || || || || ||
QY 747 cgcgcgtgtgagcgtgtgtgcgacccggaacagctgcgacgttcgcgcgcgcgga 806
   || || || || || || || || || || || || || || || || || || || || || || ||
Db 771 GAGCTGTGCTAGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 830
   || || || || || || || || || || || || || || || || || || || || || || ||
QY 807 cctgtgtgcccagcgctgcgaagagtgctgtcggtacgaacccgttcggtcagtc--aa 863
   || || || || || || || || || || || || || || || || || || || || || || ||
Db 831 CTTGTGTCCCAAGCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 890
   || || || || || || || || || || || || || || || || || || || || || || ||
QY 864 caccgcggaagctggaagtgagtgagtgagtgagtgagtgagtgagtgagtgagtgag 923
   || || || || || || || || || || || || || || || || || || || || || || ||
Db 891 GGGCGGTCGCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 950
   || || || || || || || || || || || || || || || || || || || || || || ||
QY 924 cgtgtgtgtcctgtgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 983
   || || || || || || || || || || || || || || || || || || || || || || ||
Db 951 CTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1010
   || || || || || || || || || || || || || || || || || || || || || || ||
QY 984 ttccagatcgagcggaatccggtlcccggtcgttccttcgcgcgcgcgcgcgcgcgcgc 1043
   || || || || || || || || || || || || || || || || || || || || || || ||
Db 1011 GGTGTGAGTGTGAGCGGACCGCGCGCATTCGCTTGTGGCGACGAGCATGACAGTGT 1070
   || || || || || || || || || || || || || || || || || || || || || || ||
QY 1044 cctcgtgttactactgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1100
   || || || || || || || || || || || || || || || || || || || || || || ||
Db 1071 CTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1127
   || || || || || || || || || || || || || || || || || || || || || || ||
```

```
RESULT 15
US-09-385-028-1
; Sequence 1, Application US/09385028
```

```
Patent No. 6232106
GENERAL INFORMATION:
APPLICANT: Susan E. Jensen
APPLICANT: Kwamena A. Aidoo
APPLICANT: Ashish S. Parthkar
TITLE OF INVENTION: DNA Sequence Encoding Enzymes of Clavulanic
Patent No. 6232106
TITLE OF INVENTION: Acid Biosynthesis
NUMBER OF SEQUENCES: 25
CORRESPONDENCE ADDRESS:
ADDRESSEE: JACOBSON, PRICE, HOLMAN & STERN, PLLC
STREET: The Jennifer Building, 400 Seventh Street, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30 (BPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/385,028
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/790,462
FILING DATE: 29-JAN-1997
ATTORNEY/AGENT INFORMATION:
NAME: D. Douglas Price
REGISTRATION NUMBER: 24,514
REFERENCE/DOCKET NUMBER: 1418/P57452052
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 638-6666
TELEFAX: (202) 39305350
TELEX: RCA 248593 IDEA UR
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 15079 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ORIGINAL SOURCE:
ORGANISM: Streptomyces clavuligerus
US-09-385-028-1
```

```
Query Match 11.4%; Score 137.4; DB 4; Length 15079;
Best Local Similarity 50.4%; Pred. No. 1.1e-14;
Matches 513; Conservative 0; Mismatches 471; Indels 33; Gaps 6;
QY 99 gcaacgctggcggccacgggtctgtgtaacgtctccgcttcgcggaagctgctgcggtgtct 158
Db 13897 GCTGTGGAGAGGACGACGAGGTGTGGTGTGACCTGCACGCGCGGGCGCGGCGTCT 13956
QY 159 cggagcgaagcagcttcgc-----cgagcggcatcgcgcggttctgtgacccagctgt 212
Db 13957 GGGCGACCGCGCTTCACCGCGGTGACGAGCGCGCGCGGCTTCGCGATGTGACCGCAC 14016
QY 213 cgggcccggggtgtcgcgagatcgtcggaacatcatcctgttccagagcggccgca 272
Db 14017 CTCCTCACTGTGTGGCGCAACCGGAGTGGCGTCTTCATCCGATGAGACCCGCA 14076
QY 273 ccacggcgcggtgctgcgggggtgtgtcggccgg--cgcttcgcgctccgcgctgcgcg 331
Db 14077 GCATCTCCGGCTCGCTCATGTGTACCCGGGACTTCCTGGCCCGCGCGGAGCGCT 14136
QY 332 tggaaacggtgatcgccggcaac--gtgaaagacctgtgcggcccgccctggccggag 389
Db 14137 GCGCCCGCGGCTCGGGAGCTGCTGACGAGATCTGGGCGGCTGTGTGAAGGGGAGCG 14196
```



```

/clone="RP43-040F09.TJ"
/sex="male"
/cell_type="lymphocytes"
/clone_lib="RPCI-43 Chimpanzee Male BAC Library"
BASE COUNT      0 a 778 c 655 g 0 t 365 others
ORIGIN

```

```

Query Match      9.1%; Score 109.6; DB 12; Length 1798;
Best Local Similarity 39.4%; Pred. No. 0.00014;
Matches 445; Conservative 0; Mismatches 677; Indels 7; Gaps 2;

```

```

QY 84 gggagcaagaccggtgacgtggtgcgcgcacccgggtctgtgtctccgcttcgagca 143
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1313 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1254

QY 144 cgtgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 203
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1253 GNGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1194

QY 204 cgaacctgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 263
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1193 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1134

QY 264 cgaagccgaacacagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 323
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1133 NGNGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1074

QY 324 ggcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 383
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1073 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1014

QY 384 ccggggcgcgatgacgtgtgtcgaagacgtgtacccgcgcgcgcgcgcgcgcgcgcgc 443
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1013 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 954

QY 444 cggcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 503
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 953 CGNGNGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 894

QY 504 gggagcgaacccctgacccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 558
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 893 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 834

QY 559 -gcgacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 617
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 833 CGNGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 774

QY 618 cggcgaagacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 677
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 773 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 714

QY 678 cggagatcgttaagacggtgtgtacgttcatctt-caacgcgcgcgcgcgcgcgcgc 736
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 713 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 654

QY 737 aggtggcaagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 796
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 653 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 594

QY 797 gccggcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 856
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 593 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 534

QY 857 agtccaacacccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 916
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 533 NNNGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 474

QY 917 agcaagcgtggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 976
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 473 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 414

QY 977 ccgagacatttcgacatcgaagcggatccggtccggttcgattcgttcgcgcgcgcgc 1036
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

DB 413 GNGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 354
QY 1037 gctactgccttcggtcttactactcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1096
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 353 GCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 294

QY 1097 ccgacgtccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1156
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 293 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 234

QY 1157 ccatgttcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1205
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 233 GCGNGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 185

```

```

RESULT 2
AG060116/c 1278 bp DNA linear GSS 03-NOV-2001
DEFINITION Pan troglodytes DNA, clone: PTB-047106.F, genomic survey sequence.
ACCESSION AG060116
VERSION AG060116.1 GI:16611346
KEYWORDS GSS: GSS (genome survey sequence).
SOURCE Pan troglodytes male lymphoblast DNA, clone_lib:PTB Chimpanzee Male
BAC library clone:PTB-047106.F.
ORGANISM

```

```

REFERENCE
AUTHORS Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
Totoke,Y., Watanabe,H. and Sakaki,Y.
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Pan.
1 (sites)
TITLE BAC end sequences of library PTB
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1278)
AUTHORS Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
Totoke,Y., Watanabe,H. and Sakaki,Y.

```

```

TITLE Direct Submission
JOURNAL Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC),
1-7-22 Suehiro-chou,Tsukumi-Ku, Yokohama, Kanagawa 230-0045, Japan
(E-mail:shimbe@gsr.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT Clones are derived from the chimpanzee BAC library PTB This BAC end
was generated during the R&D process and may have higher chance of
clone tracking errors.
PRIMERS
Sequencing: -21M13
LIBRARY
Vector : pKS145
R.Site 1 : SacI
R.Site 2 : SacI.
Location/Qualifiers
1..1278
/organism="Pan troglodytes"
/db_xref="taxon:9598"
/clone="PTB-047106.F"

```

```

FEATURES
source
1..1278
/sex="male"
/cell_type="lymphoblast"
/clone_lib="PTB Chimpanzee Male BAC Library"
BASE COUNT 49 a 539 c 533 g 33 t 124 others
ORIGIN

```

```

Query Match      8.7%; Score 104.8; DB 12; Length 1278;
Best Local Similarity 46.7%; Pred. No. 0.00053;
Matches 397; Conservative 0; Mismatches 441; Indels 13; Gaps 5;

```

```

QY 128 tctccgcgttcgagcgtgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 187
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1257 TCTGCTGTGTCGNTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1198

QY 188 tccgcgcgttcgagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 247
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

[illegible][illegible]

Oy 742 gggaaacgcctgtctgtagctctgtgtgcagacacggagacacagctcgtactctctccgagccgg 801
 Db 635 gggggcggggggccgcccnnnnnccggccggccggccggccggccggccggccggccggccggccgg 576
 Oy 802 ccgagactgtctgtgtcccaagcgcgttcgagagatgtgtctgtgtgtacagaccgctgtgtcagctc 860
 Db 575 gcgcgcnnnccggcggggcggccnccggccggccggccggccggccggccggccggccggccggccgg 517

LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM
AG032943/c	AG032943	Pan troglodytes DNA, clone: PTB-007N10.F, genomic survey sequence.	1542 bp	DNA	linear	GSS 01-NOV-2001
	AG032943	AG032943.1 GI:16559816		GSS; GSS (genome survey sequence).	Pan troglodytes male lymphoblast DNA, clone_1lb:PTB Chimpanzee Male	
		BAC library clone:PTB-007N10.F.				
		Pan troglodytes				

REFERENCE AUTHORS	TITLE	JOURNAL REFERENCE AUTHORS
1 (sites) Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T. D., Yada, T., Tokoki, Y., Watanabe, H. and Sekaki, Y.	BAC end sequences of library PTB	
Unpublished		
2 (bases 1 to 1542) Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T. D., Yada, T., Tokoki, Y., Watanabe, H. and Sekaki, Y.		

TITLE Direct Submission
JOURNAL Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC): 1-7-22 Suenho-cho, Tsukumi-Ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: chinpbes@gsc.riken.go.jp, URL: <http://hgp.gsc.riken.go.jp/> Tel: 81-45-503-9111, Fax: 81-45-503-9170)
COMMENT Clones are derived from the chimpanzee BAC library PRB. This BAC end was generated during the R&D process and may have higher chance of clone tracking errors.
PRIMERS

```

Sequencing: -21M13
LIBRARY
Vector      : pKS145
R.Site 1    : SacI
R.Site 2    : SacI.
Location/Qualifiers
1. .1542
/organism="Pan troglodytes"
/db_xref="taxon:9598"
/clone="PTB-007M10.F"
/sex="male"
/cell_type="Lymphoblast"
/clone_id="PTB Chimpanzee Male BAC Library"
BASE COUNT
ORIGIN
61 a      657 c      614 g      19 t      191 others

```

Query Match	8.4%	Score 101.8	DB 12	Length 1542
Best Locality	45.9%	Pred. No. 0.0012		
Matches 512	Conservative	0	Mismatches 590	Indels 14
				Gaps 8

[illegible][illegible]

RESULT	5
LOCUS	BG809984/c
DEFINITION	BG809984
ACCESSION	mctc002xdllf Magnaporthe grisea Appressorium Stage
VERSION	grisea cDNA clone mgtc002xdllf 5' mRNA sequence.
KEYWORDS	BG809984 GI:14180964
SOURCE	EST. Magnaporthe grisea.

	Query Match	8.3%;	Score 100.6;	DB 10;	Length 1516;
	Best Local Similarity	43.2%;	Pred. No. 0.0016;		
	Matches 489;	Conservative	0;	Mismatches 624;	Indels 19; Gaps
Oy	58 cccataaccggtccctacacacttgatctgtgcgcgaagcccgcgtgcaacagtgtgcgccacacgg	117 			
Dd	107 CCTCCTCCGGGAGNAACTTTTCNCCCCCCCCCAAAACCNNTTGNMNGSGGGGGCGNNNCAC	166 			
Oy	118 gctcgtgaagtcctcccgcttcgcgaagctgtgcgcgtgtgtctcgcgcgaagagctctgcgc	177 			
Dd	167 CCNNNGMCNCNCNCCGCGGGGAGNGGCCCGGCNCGCCGGGGCCGCGGGCCGGGAGACANCTG	226 			
Oy	178 cgagccgcgcatccgcgcgtttcttgcagccacctcgtgcgcgcgcgcgcgtctgcgcgagatc	237 			
Dd	227 TCTCCCGCGGGGCGCGCGCCGCCGCCGCCGCMNGCGCGCCGCCGCCGCCGCGCGCGCGCCG	286 			
Oy	238 gtccgcgacaatcatcttgttccaagaaggaccgaacccgcgcgcgcgtgcgcggtgtgtc	297 			
Dd	287 CCGCGCCGCGGGCGCGCGGCCGCCGCCGCCGCMNGCGCGCCGCCGCCGCCGCGCGCGCGCCG	346 			
Oy	298 ggccccgcgttctcgcgcgttcgcgcgtgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc	357 			
Dd	347 CCGGGCCCGCGCGCGCGGGCGCGGCCGCCGCCGCCGCCGANCNCCGCMCGCCNCCNMGCGCGGC	406 			
Oy	358 gaacacattctgtgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc	414 			
Dd	407 CGCGCGCGCGCMNGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC	466 			
Oy	415 gcgtacccgcgtgcgcgtgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc	474 			
Dd	467 GCGCGCCCGCGCGCGCGCGCGCGCGCMNGMNGMNGGAGNBGGCGCGCGGGGGGGGGCGCMNC	526 			
Oy	475 ggagcggtctgcgcgtctgtcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc	534 			
Dd	527 NGCMCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC	586 			
Oy	535 gaagacatgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc	594 			
Dd	587 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCN	646 			
Oy	595 ctgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc	654 			
Dd	647 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCMNGCGCGCGCGCGCGC	706 			
Oy	655 gaaccgcgcgtgatgatgcgaagaagatcgtacgaagctgtatcaagtatcatctaac	714 			
Dd	707 CGCGG-GCGCGCGCGCGCGCGCGCGCGCGCMNGCGCMNMNGGGGCCCCCMCGCGCGCGCGCGC	765 			
Oy	715 ggccacagagaaagctgtgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc	774 			
Dd	766 GGNGCGCGCGCGCGCGCGCMGGCCGCGCGCGCMGCGCGCGCGCGCGCGCGCGCGCGCGCMNGC	825 			
Oy	775 gaacacatctgcacctgtctcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc	834 			
Dd	826 GCGCGAAGCGCGCGCGCGGGGCGGGGCGGCMGCGCGCGCGCGCGCGCGCGCMGGCGCGG	885 			
Oy	835 ctgcgcgtacacccgcgtctgtgcagttccaacaaccgcgcgcgcgcgcgcgcgcgcgcgcgcgc	894 			
Dd	886 GCGCGCGCGCGCGCGCGG---CGCGCGCGCMCMGCMGCGCGCGCGCGCGCGCGCGCGCGG	941 			
Oy	895 cgagctgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc	954 			
Dd	942 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	1001 			
Oy	955 gaaccgcgcgcgtacgaagacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc	1014 			
Dd	1002 GCGCGCGCGCMNGCGCGCGCGCGCGAGAAGCGCMCGACGCGCGCGCGCGCGCGCGCGCGCGC	1061 			
Oy	1015 atgtctctcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc	1063 			
Dd	1062 CGCGCGCGCGCGCGCGCGCGCMNGMNGCGCGCGCGCGCGCGCGCGCMGCGCGCGCGCGCGC	1121 			
Oy	1064 gtacgcgcgtcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc	1123 			

[illegible]

[illegible][illegible]

[illegible][illegible]

.


```
gene
5390..6085
/gene="biod"
CDS
5390..6085
/gene="biod"
/function="biotin pathway"
/codon_start=1
/transl_table=11
/evidence=experimental
/product="desthiobiotin synthetase"
/protein_id="AAB17460.1"
/db_xref="GI:1277028"
/translacion="MRGFVGTDEVKTVISSGLAALIKDNNRHVYKPKLSGIS
RHHDSPTSLKDMSPSLSHEDITPPFAFKAPLAPYAGKLEGTVMEEVLSHMGRI
REKHECFIVGAGGISVPLGEDYLVSHVKAQLDPMIVARPLGTINTFTLVKRAE
SMGLPIGIIINGISDSPEDEKTNPEMERLQGVPLGVTPKLANVTKETVLAHMYKD
HINLSILMNOGV"
6088..7095
/gene="biod"
6088..7095
/gene="biod"
/function="biotin pathway"
/codon_start=1
/transl_table=11
/evidence=experimental
/product="cytochrome P450-1 like enzyme"
/protein_id="AAB17462.1"
/db_xref="GI:1277030"
/translacion="MTIASSSTASEFLKNPYSFYDPLRAVHPITYKSPLYKPVYVTS
VEETAAILDKARPKVPTPLPESTKYODLSHONOMMLFONOPDRRLRTLASGATP
RTTESYQPYLTFTVHLLDNOGKKMEVTSFAPPLASVIANITIGPEEDREOLKE
MAASLIQTDITFRSKALTEGNTMAVOAMAYRKELOKRRHPQDMISMLKREKED
KITEPASTGICILAIAGHETTVNLISNSVLCILQHPEDLIRENPDLIGTAVSECL
RRESPOMTARVASIEDICGVITRQGEQVYLLGAANDPSIFTPNPVDFDITRPNP
HLSFGHHVCLGSSILARLEAOIAINTLLQRMPSINLADFEWRYRPLGFRALDELPA
TPE"
8428..9189
/gene="orf2"
8428..9189
/gene="orf2"
/function="unknown"
/codon_start=1
/transl_table=11
/product="unknown, probably a redox enzyme"
/protein_id="AAB17463.1"
/db_xref="GI:1277031"
/translacion="MKKVLIAAGNGVIGRLIAEGLISDYEVTVLDKHFHDKASSIOA
DAANEELKRIPKDIDAILNLAVIKYDIDIAEFKMDVDFYRASYLCRAAEL
GIQKLVFASNHVTVYKERGSLRGREITTSDDYLSKRLGVLTSEOGIHLFYLE
NKLIVINLRIGTVVDEMDTLHEKRETKTLLSHDLSIFKAALETNIRKGTYYAVS
DNRPWSTESVNLGFSPOINTAELNEENGA"
9375..10153
/gene="orf3"
9375..10153
/gene="orf3"
/codon_start=1
/transl_table=11
/product="unknown, probably a redox enzyme"
/protein_id="AAB17464.1"
/db_xref="GI:1277032"
/translacion="MGATASHLSKRLPDLIYGFLLMALICVLPFTIVIASPFAVE
EYVSKFTILPTTFSLDARYRYIESPDIIKSLVSEFVTVGTAVSMFLSLMAYGLS
RDDLIGROPIMFLVFTFMFSGGMIFELVYKSLDLSVWALILPTAINFNLILK
NFOQNPSSLESARKIDGNDIGIFKIVLPISLPAIATISLFAVAVTWNTRYMTAILY
LNDSAKWPLOVLLROIYIVYSGMGDMSEMGSGSPPEOTIG"
```

```
BASE COUNT      2886 a      2093 c      2502 g      2672 t
ORIGIN
alignment_scores:
  Quality: 609.00      Length: 396
  Ratio: 2.316      Gaps: 5
  Percent Similarity: 66.414      Percent Identity: 33.586
alignment_block:
US-09-724-797-36 x BSU51868 ..
Align seg 1/1 to: BSU51868 from: 1 to: 10153
8 AlphaheaprophrAspAlaaspValArgArgaspprotyrProserTy 24
||| .....
7173 GCATCGCAACGTCATCTTCTGAGTTTGTGAAAACCATATTCCTTTTA 7222
24 rHsTrpLeuLeuArgHisaspproValHisArgGlyAla.....H 38
||| .....
7223 CGACACATTCGAGAGCTGTCATCCATTCATTAAGGAGCTTCTTAAAT 7272
38 IsArgValITrpyrValaserArgpheaLaspyalarGalaValLeuGly 54
::: ||||| .....
7273 ACCCGGCTGTATGTCACAGATATGMAAGAAAGCGCTGATATTGAAA 7322
55 AspGluArgPhe...AlaArgThrGlyLeuArgArgPheThrAspLe 70
||| ||||| .....
7323 GATGCGAGATTCAAGTCCGCCACCCCGCTGCCGAGACGTCAACCAATA 7372
70 uValGlyProGlyLeuLeuAlaGluIleValGlyAspIleLeuPheG 87
||| ||||| .....
7373 TCAGGAC.....CTTCAACATGTGCMAAATCAAGATGCTGTTTC 7413
87 lnsAspGluProAspHisGlyArgLeuArgGlyAlaValGlyProAlaPhe 103
||| ||||| .....
7414 AGAACCGAGCTGATCATAGCATGTCGAGCGCTGCCAGCGAGGCTTT 7463
104 SerProSerAlaLeuArgArgLeuGluProValIleAlaGlyThrValAs 120
::: ||| .....
7464 ACCCGGAGAACGACAGAGATTATCAGCGGTATATCATTTGAAACGTCCA 7513
120 pAspLeuLeuArgProAlaLeuAlaArgGlyAlaMetAspValValAspG 137
||| ||||| .....
7514 TCATTTGCTTGATCAAGTCACAGTGAATAAAGAGTGGCTCATTTCCG 7563
137 lueuAlaTyrrProLeuAlaLeuArgAlaValLeuGlyLeuLeuGlyLeu 153
::: ||||| .....
7564 ACTTGCCTTTCTCTTAGCAGATTGTGTCATACCTTAACATTAGGTGA 7613
154 ProAlaAlaAspTrpAlaValAlaGlyArgTrpSerArgAspValGlyAr 170
||| ||| .....
7614 CCGAGGAGAGATAGGAGCAATTAAAGAGTGGCTGCAGACTCATTTCA 7663
170 gThrLeuAspArgGlyAlaSerAlaGluSpmelaArgArgGlyHisAlaA 187
::: ||||| .....
7664 AACGATTGATTTTACCCTCAAGAAAGCATTAACAGGGCAATATTA 7713
187 lAlleAlaGluPheaLaaspyrValGluArgAlaLeuAlaArgArgArg 203
::: ||| .....
7714 TGGCTGTGAGGCTATGCGCATATTCAAGAGCTGATTCMAAAGGAAA 7763
204 ArgGluGlyGlyLusPheLeuLeuAlaLeuMetLeuAspAlaHisAspAr 220
||| ||||| .....
```

```

7764 CGCCACCCCTCAACAGATATGATGATGATGCTTTGAAGGGAGAGAAAA 7813
220 ggilyleuMetSerArgAsnGluIleValSerThrValValThrPheLeu 237
7814 GGATTAAGCTCAACGAAGAGAGCGGCGCATCTTTCGGCATATGTCGGCGCA 7863
237 heThrGlyHisGluThrValAlaSerGlnValGlyAsnAlaValLeuSer 253
7864 TCGCCGGACATGAGAACACGTCATCTCACTCAGACAATTCAGTCTTGT 7913
254 LeuLeuAlaHisProAspGlnLeuAspLeuLeuArgArgProAspLe 270
7914 CTGCTGAGCATTCAGAACAGCTTTTGAACCTGAGAGAAAAATCCAGATCT 7963
270 uLeuAlaGlnAlaValGluGluCysLeuArgTyrAspProSerValGlns 287
7964 TATTGGACCGCGCATGAGGAATGTTTACGCTATGAAGCCGCCAGCAA 8013
287 eArgnThrArgGlnLeuAspValAspValGluLeuArgGlyArgArgLeu 303
8014 TGACAGCCACAGATTCGCTCAGAGATATTGACATCTCGCGGTCAGCATC 8063
304 ArgArgAspAspValValValValLeuAlaGlyAlaAlaAsnArgAsp 320
8064 CGCTAACGAGAACACAGCTATCTTTTGTTAGAGCGGCTTAATCCAGACC 8113
320 cArgArgTyrAspArgProAspAspPheAspIleGluArgAspProValP 337
8114 TAGCATATTCACGACACCCGATGCTTTCGATATTCAGAGAGTCTTAATC 8163
337 roSerMetSerPheGlyAlaGlyMetArgTyrCysLeuGlySerTyrLeu 353
8164 CGCATCTTTTCATTTCGGGCATGGCCATCATGTTGCTTGAAGGTCCTCGGTG 8213
354 AlaArgThrGlnLeuArgAlaAlaVal...AlaAlaLeuAlaArgLeuPr 369
8214 GCAGCATTTAGACGGCAAAATTCGATTAACACTCTTCGACAGCGAATGCC 8263
369 cGlyLeuArgLeuGlyCysAlaSerAspAlaLeuAlaTyrGlnProArgT 386
8264 CAGCCTTAATCTT....CGCGATTTCGAATGGCGGATCGGCGGCTTT 8307
386 hrMetPheArgGlyLeuAlaSerLeuProIleAlaPhe 398
8308 TTGGATTTCGGCGCTTGAGAGACTTCGCGGTACTTTT 8345
seq_name: gb_ba:BSUB0016

```

self document	Accession	215540 bp	DNA	linear	Dec 26-NOV-1997
LOCUS	BSUB00016				
DEFINITION	<i>Bacillus subtilis</i> complete genome (section 16 of 21) : from 2997771 to 3213410.				
ACCESSION	299119 AL009126				
VERSION	299119.1	GI:2635411			
KEYWORDS					
SOURCE	<i>Bacillus subtilis</i> .				
ORGANISM	<i>Bacillus subtilis</i> .				
REFERENCE	<i>Bacteria</i> , Firmicutes; <i>Bacillus</i> / <i>Clostridium</i> group; <i>Bacillus</i> / <i>Staphylococcus</i> group; <i>Bacillus</i> .				
AUTHORS	1 (bases 1 to 215640) Kunst, F., Ogasawara, N., Moszer, I., Albertini, A.M., Alloni, G., Azevedo, V., Bartorel, M.G., Bessieres, P., Bolotin, A., Borchert, S., Boriss, R., Boursier, L., Brans, A., Braun, M., Brignell, S.C., Bron, S., Brouillet, S., Brusch, C.V., Caldwell, B., Capano, V., Carter, N.M., Choi, S.K., Codani, J.J., Conerton, I.F., Cummings, N.J., Daniel, R.A., Denizot, F., Devine, K.M., Dusterhoef, A., Ehrlich, S.D., Emmerse, P.T., Estlin, K.D., Errington, J., Fabel, C., Ferrari, E., Foulger, D., Fritz, C., Fujita, M., Fujita, Y., Fuma, S., Gallizzi, A., Gallardo, N., Gauseppl, G., Guy, B.J., Haga, K., Halech, Y., Goltightly, E.J., Grandi, G., Giuseppe, G., Guy, B.J., Haga, K., Halech, Y., Harwood, C.R., Henaut, A., Hilbert, H., Holsappel, S., Hosono, S., Hullo, M.F., Itaya, M., Jones, L., Joris, B., Karamata, D., Kasahara, Y., Klaerr-Blanchard, M., Klein, C., Kobayashi, Y., Koetter, P.,				

FEATURES	TITLE
source	The complete genome sequence of the gram-positive bacterium <i>Bacillus subtilis</i>
gene	1. 215640
terminator	location/Qualifiers
CDS	organism="Bacillus subtilis"
	/strain="168"
	/db_xref="taxon:1423"
	complement(8..1001)
	/gene="ytmn"
	complement(8..24)
	/gene="ytmn"
	complement(99..1001)
	/gene="ytmn"
	/function="unknown"
	/codon_start=1
	/transl_table=1
	/protein_id="CAB14906.1"
	/db_xref="GI:2635412"
	/db_xref="SPTRMBL:034430"
	/translation="MKRLVPAFIGLSLIDSGMAYGVTSLSLLARCTPAVAS
	ASVIAEYVITTAAGSVSHIRKGNVDKQTVQVLTPGSIAGALGAFTLSQDPVAKPP
	ISPLLIGVYLIRFLFQYKPALEKHPVLRNROSIPGLVIAFDATGGGNGPVAT
	TPILSRKSGSPKRVGVGTSEPAIVASATAGLISLGMGVMLVWFSLMAGGI
	APTAIVLQKRFHPOLMGVINGCFITLVNARLLINEMVIANVHPLITYAIGAIWLSAV
	LPLSKTGNRNVTAADVHAKER
	complement(1020..2270)
	/gene="hipo"
	complement(1020..2270)
	/gene="hipo"
	/EC_number="3.5.1.32"
	/codon_start=1
	/transl_table=1
	/product="hipurate hydrolase"
	/protein_id="CAB14907.1"
	/db_xref="GI:2635413"
	/db_xref="SPTRMBL:0344980"
	/translation="MSLDYRNINIGSPYQVTGNDITLKEESNPVNLTKELQIGI
	RRHLHQYBELSKKEFEETTAFTIKCLKEKGIDIRTPALKTGVFADIAGSEBPALARR
	DIDLDPILEKTEGLPYASKHGMHACGHDFHTAALLGAFTLKEKODSLKGRIRLFE
	PAEAGAGATVIEDGQDGDIDAVIGLHNKDPDIAGVLTGPIAMAADVRFVIEE
	KGAIAAPHNGFDPLIGASOLIVALTQTSRNVNPLQSAITLGTGKINGSTWNVIPDGI
	VVTCGYATPDSEVNRQKQKQRFVTEQISAFSLKAWKHSQPPRCNDPAITGLVLA
	RDAAHRAKLVQITDPAFSTAGEDPAYILEHIGSFATFTQDDHDMHNAFTIDEAL
	KASYFLYESARLILDSNEESKISD
	complement(2288..2980)
	/gene="fibr"

```
CDS      complement(2288..2980)
          /gene="ribB"
          /function="regulation of riboflavin biosynthesis genes"
          /note="alternate gene name: ytnK"
          /codon_start=1
          /transl_table=11
          /product="transcriptional regulator"
          /protein_id="CAB14908.1"
          /db_xref="GI:2635414"
          /db_xref="SMISS-PROT:P94465"
          /translation="MTIAGTVKKGKQGRKGFPTANYDAKIHGLRNGVYGLATVN
          HQFHLDGVNMGVAPGVSNLEKTLLEIFEDPHRDYGEKIECSILFKIIEERFPSLE
          SLTKQIKDIDSCVAKRFELIGIMAPRKESLSHQLNPDLCFYKKCNMLYGVRGV
          YVINDMVFEEYGTQVAYRIRYILSFLSKEDNPKVSSKYIRFGAGGLADKLNFIS
          SYVESEENILC"
          complement(3027..4355)
          /gene="ytnJ"
          complement(3027..4355)
          /function="unknown"
          /note="similar to nitrlotriacetate monooxygenase"
          /codon_start=1
          /transl_table=11
          /protein_id="CAB14909.1"
          /db_xref="GI:2635415"
          /db_xref="SPTREMBL:O34974"
          /translation="MTRADFIQGAMIHGVGGTTPDGMRRHDPVDSASTNIEPYMKAO
          TAEKGLFSFIADGLFISEKISPHPLINPEPITLISALASTKNIGLVGTSTSTPTE
          PPTISROLMSLDIHISGGRGRNMLVTSPOGARNNKSKMLPHTERTETLAQCHLDVVR
          GLMNSWEHDAFTHNKKTGFQPOAKLRLNHKGYEQVAGPLNIGRSKQGEFVYQNG
          SSETGQFPAKNADAIEPHNSNLEETKAFYADVKSRADEGDPSSVRIFPQISPIVA
          DTEBEAKKYREFAEILIPLENAVLYLAFEDDYDLISVPLDEPFDIGGVNAPQST
          DTRIKREAKARNLTREVAQEMAFPRTEIGCPENVASIETWFMAPADGIVASDI
          PGTLAFVKEKVIPIIOERGLYRQDYGRTGLRNLGLGIPQHSVLSHSH"
          complement(4352..4633)
          /gene="ytnI"
          complement(4352..4633)
          /function="unknown"
          /codon_start=1
          /transl_table=11
          /protein_id="CAB14910.1"
          /db_xref="GI:2635416"
          /db_xref="SPTREMBL:O34639"
          /translation="MSDVNIYVMSKKGCSYCEEVKNYLNENKFPONIDYSEKELR
          DILQVYGVRAHYPVVEIGRNOYOGITEIGIHLALANHAQIKEAR"
          complement(4648..5652)
          /gene="ytnO"
          complement(4648..5652)
          /function="unknown"
          /note="similar to hypothetical proteins"
          /codon_start=1
          /transl_table=11
          /protein_id="CAB14911.1"
          /db_xref="GI:2635417"
          /db_xref="SPTREMBL:O34846"
          /translation="MIRLSILDPSLIGGETAADLTQHTVKVLAQMAEEGCGYHREWVAE
          HNNDEIGSAPVELLYLAASTRTKIRIAGSVMLCHYSVYKVAQOPHILSALAPRT
          DLGVKAPGSGROLSTDALQAEKKVPYROPDELELTFRPDDPFDITRIYALAPREO
          VDRKGCITLIGSTESAISAALGTSVFVAFINDEEVLLEEARAFDAHLPPGSEAR
          PHLAFAVPAHKTKEAEKHIYVESIKYVLDGKRKVVNVSREDAVAYLENTPEYDII
          YOKTGIAGTKEEVAEELRLSGTYKINDFYIFPIKNAVEKQISYQLSDAVLAAR"
          complement(5649..6428)
          /gene="hisP"
          complement(5649..6428)
          /gene="hisP"
          /codon_start=1
          /transl_table=11
          /product="histidine transport protein (ATP-binding
          protein)"
          /protein_id="CAB14912.1"
```

```
/db_xref="GI:2635418"
/db_xref="SPTREMBL:O34900"
/translation="WIEIKNTHKQFGJHNLKGINLTVRGEVVTIIGPSGSGKTFPL
RCLNLERPDDEGIISIDHKVINCPSKSEVHMLRKOTAMFOQYHLPARKVTIENY
EGITIAKMRKQDAVAAVAENELRKVQDLSODKSNVSSGSGKQGVIAIALIHPIV
LIPDEPTALDPELVGEYLEMTEIVTGATMIVYTEMPEARVSDQVVFMEGCVIV
EOGTPPEVFRRTKKDRTPQFLRRVSPELLEPKHKEPY"
complement(6425..7132)
/gene="ytnM"
complement(6425..7132)
/gene="ytnM"
complement(7162..7881)
/function="unknown"
/note="similar to amino acid ABC transporter (permease)"
/codon_start=1
/transl_table=11
/protein_id="CAB14913.1"
/db_xref="GI:2635419"
/db_xref="SPTREMBL:O34931"
/translation="MORDPEPPIVAMKEVMKTPITLMAVLPVREGVYALGNITVR
IFRKGLVACSRFTYVSFFRSPALHMLIYLGIPIVADOVSSFFHLGMSANEIPVSM
FVIALSLTAGAVLEIRSGITLMDTGQVAAVSIQTVSOTFRVILQALKVSIIP
NFTMLIGFLHTTSIAAIVAVPEITGTATVASDNVAFLEAFIGALITVYVTLILIES
ANGVLERKAARFOCGTL"
complement(7162..7881)
/gene="ytnL"
complement(7162..7881)
/gene="ytnL"
/function="unknown"
/note="similar to amino acid ABC transporter (permease)"
/codon_start=1
```

```
alignment_scores:
  Quality: 609.00      Length: 396
  Ratio: 2.316        Gaps: 5
  Percent Similarity: 66.414  Percent Identity: 33.586
```

alignment_block:

US-09-724-797-36 x BSUB0016/rev ..

Align seg 1/1 to reverse of: BSUB0016 from: 1 to: 215640

```
8  AAlpheasProthraspalaaspValaArgargProtyrProserTy 24
   |||  ::  ::::::::::::::  ::::::::::::::  ||
  91715 GCATCGTCAACGTCATCTGTGAGTTTGAAAACCCATATCTTTTAA 91666

24  rHtstPLeuLeuAArgHisAspProValHisArgGlyAla.....H 38
   |||  ::::::::::::::  ::::::::::::::  ::
  91665 CGACACATTGCGAGCTTCATCCCTATTAAGGAGCTTCTTAAT 91616

38  iSArgValTrpTyValserArgpheAlaaspValaArgAlaValleuGly 54
   ::  |||  |||  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::
  91615 ACCCGGCGTGTGTCACACAGATATGAAGAAACGGCTCTATTTTGAAA 91566

55  AspgluArgphe...AlaArgThrGlyLeaArgargPheTrpThrAsple 70
   |||  |||  |||  ::  ::  |||  |||  ::  ::  ::  ::  |||
  91565 GATGCGAGATTCATAAGTCCGACCCCGCTGCTGAGAGCTCAACCAATA 91516

70  vValGlyProGlyLeuLeuAlaGluIleValGlyAspIlelleuPheG 87
   |||  ::::::::::  ::::::::::  ::::::::::  ::::::::::  |||
  91515 TCAGGAC.....CTTTCACATGTGCMAAATCAAAAGATGCGTTC 91475

87  lAspGluProAspHisGlyArgLeuArgGlyValaValGlyProAlaPhe 103
   ::  ::  |||  |||  |||  |||  |||  |||  |||  |||  |||
  91474 AGAACAGCCTGATCATATACGATTCGCGACGCTGCCAGCGAGCGTTT 91425

104  SerProSerAlaLeuArgArgLeuGluProValIleValaglyThrValas 120
   ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::
  91424 ACGCGAGAACGACAGAGATTATCAGCGGTATATCATGGAATGTCOCA 91375

120  pAspLeuLeuArgProAlaLeuAlaArgGlyAlaIleAspValaValaspG 137
   |||  |||  |||  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::
  91374 TCATTTGCTTGATCAAGTCAAGGTAAAMAAAGATGAGGTCATTTGGG 91325
```

[illegible]

234-281
90234-90257

seq_name:	gb_ba:AF008320		
seq_documentation_block:			
LOCUS	AF008320	220060 bp	DNA linear BCT 04-FEB-1998
DEFINITION	Bacillus subtilis rrmB-dnaB genomic region.		
ACCESSION	AF008320		
VERSION	AF008320.1	GI:2293135	
KEYWORDS			
SOURCE			
	Bacillus subtilis.		

ORGANISM	REFERENCE
Bacillus subtilis	Bacteria: Firmicutes: Bacillus/Clostridium group:
Bacillus/Staphylococcus group: Bacillus.	
1 (bases 1 to 7430)	
Green,C.J., Stewart,G.C., Hollis,M.A., Vold,B.S. and Bott,K.F.	
Nucleotide sequence of the Bacillus subtilis ribosomal RNA operon, rrb	
Gene 37 (1-3), 261-266 (1985)	
86031361	
2 (bases 153210 to 153762)	
Comors,M.J., Mason,J.M. and Setlow,P.	
Cloning and nucleotide sequencing of genes for three small,	
acid-soluble proteins from Bacillus subtilis spores	
J. Bacteriol. 166 (2), 417-425 (1986)	
86195826	
3 (bases 213161 to 218473)	
Ogasawara,N., Moriya,S., Mazza,P.G. and Yoshikawa,H.	
Nucleotide sequence and organization of dnaB gene and neighbouring	
genes on the Bacillus subtilis chromosome	
Nucleic Acids Res. 14 (24), 9989-9999 (1986)	
87117549	
4 (bases 200404 to 201481)	
Seki,T., Yoshikawa,H., Takahashi,H. and Saito,H.	
Cloning and nucleotide sequence of phoP, the regulatory gene for	
alkaline phosphatase and phosphodiesterase in Bacillus subtilis	
J. Bacteriol. 169 (7), 2913-2916 (1987)	
87250247	
5 (bases 201234 to 203212)	
Seki,T., Yoshikawa,H., Takahashi,H. and Saito,H.	
Nucleotide sequence of the Bacillus subtilis phoR gene	
J. Bacteriol. 170 (12), 5935-5938 (1988)	
89053932	
6 (bases 142232 to 144147)	
Grundy,F.J. and Henkin,T.M.	
Cloning and analysis of the Bacillus subtilis rpsd gene, encoding	
ribosomal protein S4	
J. Bacteriol. 172 (11), 6372-6379 (1990)	
91035248	
7 (bases 133624 to 134990)	
Henkin,T.M., Grundy,F.J., Nicholson,M.L. and Chambliss,G.H.	
Catabolic repression of alpha-amylase gene expression in Bacillus	
subtilis involves a trans-acting gene product homologous to the	
Escherichia coli lacI and galP repressors	
Mol. Microbiol. 5 (3), 575-584 (1991)	
91260441	
8 (bases 140810 to 142610)	
Henkin,T.M., Glass,B.L. and Grundy,F.J.	
Analysis of the Bacillus subtilis tyrS gene: conservation of a	
regulatory sequence in multiple tRNA synthetase genes	
J. Bacteriol. 174 (4), 1299-1306 (1992)	
92138624	
9 (bases 217570 to 220060)	
Putzer,H., Gendron,N. and Grunberg-Manago,M.	
Co-ordinate expression of the two threonyl-tRNA synthetase genes in	
Bacillus subtilis: control by transcriptional antitermination	
involving a conserved regulatory sequence	
EMBO J. 11 (8), 3117-3127 (1992)	
92347349	
10 (bases 134990 to 141290)	
Grundy,F.J., Waters,D.A., Takova,T.Y. and Henkin,T.M.	
Identification of genes involved in utilization of acetate and	
acetoin in Bacillus subtilis	
Mol. Microbiol. 10 (2), 259-271 (1993)	
95020526	
11 (bases 162129 to 164080)	
Grundy,F.J., Waters,D.A., Allen,S.H. and Henkin,T.M.	
Regulation of the Bacillus subtilis acetate kinase gene by CcpA	
J. Bacteriol. 175 (22), 7348-7355 (1993)	
94042910	
12 (bases 7009 to 15526)	
Kiel,T.A., Boels,J.M., Beldman,G. and Venema,G.	
Glycogen in Bacillus subtilis: molecular characterization of an	
operon encoding enzymes involved in glycogen biosynthesis and	

JOURNAL MEDLINE REFERENCE AUTHORS TITLE	degradation Moi. Microbiol. 11 (1), 203-218 (1994) 94195107 13 (bases 196487 to 200620) Jin,S. and Sonenshein,A.L. Identification of two distinct Bacillus subtilis citrate synthase genes
JOURNAL MEDLINE REFERENCE AUTHORS	J. Bacteriol. 176 (15), 4669-4679 (1994) 94321340 14 (bases 16985 to 19588) Abe,A., Koide,H., Kohn,T. and Matabe,K. A Bacillus subtilis spore coat polypeptide gene, cots Microbiology 141 (Pt 6), 1433-1442 (1995) 95400496 15 (bases 131934 to 133970) Bolotin,A., Khazak,V., Stoyanova,N., Ratnanova,K., Yomantas,Y. and Kozlov,I.
JOURNAL MEDLINE REFERENCE AUTHORS TITLE	Identical amino acid sequence of the ara(G) gene products of Bacillus subtilis 168 and B. subtilis Marburg strain Microbiology 141 (Pt 9), 2219-2222 (1995) 96118703 16 (bases 25258 to 31212) Rowland,B., Hill,K., Miller,P., Driscoll,J. and Taber,H. Structural organization of a Bacillus subtilis operon encoding menaquinone biosynthetic enzymes Gene 167 (1-2), 105-109 (1995) 96144257 17 (bases 196487 to 200620) Jin,S., De Jesus-Berrios,M. and Sonenshein,A.L. A Bacillus subtilis malate dehydrogenase gene J. Bacteriol. 178 (2), 560-563 (1996) 96134995 18 (bases 129888 to 132207) Varon,D., Brody,M.S. and Price,C.W. Bacillus subtilis operon under the dual control of the general stress transcription factor sigma B and the sporulation transcription factor sigma H Mol. Microbiol. 20 (2), 339-350 (1996) 96310371 19 (bases 81540 to 91690) Bower,S., Perkins,J.B., Vocum,R.R., Howitt,C.L., Rahaim,P. and Pero,J.
JOURNAL MEDLINE REFERENCE AUTHORS TITLE	Cloning, sequencing, and characterization of the Bacillus subtilis blotin biosynthetic operon J. Bacteriol. 178 (14), 4122-4130 (1996) 96312354 20 (bases 49093 to 51682) Vocum,R.R., Perkins,J.B., Howitt,C.L. and Pero,J. Cloning and characterization of the metE gene encoding S-adenosylmethionine synthetase from Bacillus subtilis J. Bacteriol. 178 (15), 4604-4610 (1996) 96345628 21 (bases 100760 to 102298) Kappes,R.M., Kempf,B. and Bremer,E. Three transport systems for the osmoprotectant glycine betaine operate in Bacillus subtilis: Characterization of Opud J. Bacteriol. 178 (17), 5071-5079 (1996) 9639357 22 (bases 1 to 220060) Lapidus,A., Galleron,N., Sorokin,A. and Ehrlich,S.D. Sequencing and functional annotation of the Bacillus subtilis genes in the 200 kb rnmB-dnaB region Microbiology 143 (Pt 11), 3431-3441 (1997) 96048467 23 (bases 1 to 220060) Lapidus,A., Galleron,N., Sorokin,A. and Ehrlich,D. Direct Submission Submitted (13-JUN-1997) Laboratoire de Genetique Microbienne, INRA, Domaine de Vilvert, Jouy-en-josas cedex 78352, France
JOURNAL MEDLINE REFERENCE AUTHORS TITLE	Location/Qualifiers I..220060 /organism="Bacillus subtilis" /db_xref="taxon:1423"
FEATURES	
Source	

```

RNA      485..2034
        /product="16S ribosomal RNA"
RNA      2203..5129
        /product="23S ribosomal RNA"
RNA      5185..5300
        /product="50S ribosomal RNA"
RNA      5322..5397
        /note="codon recognized: GUA"
RNA      5430..5504
        /product="tRNA-Val"
RNA      5543..5618
        /note="codon recognized: AAA"
RNA      5629..5715
        /note="codon recognized: CUG"
RNA      5721..5795
        /product="tRNA-Leu"
RNA      5810..5895
        /note="codon recognized: GGC"
RNA      5905..5981
        /product="tRNA-Gly"
RNA      5997..6073
        /note="codon recognized: UUA"
RNA      6079..6151
        /product="tRNA-Lys"
RNA      6172..6248
        /note="codon recognized: CUA"
RNA      6251..6327
        /product="tRNA-Met"
RNA      6334..6425
        /note="codon recognized: AUG"
RNA      6443..6519
        /product="tRNA-Ile"
RNA      6531..6607
        /note="codon recognized: UCA"
RNA      6607
        /product="tRNA-Ser"
RNA      6607
        /note="codon recognized: UCA"
RNA      6607
        /product="tRNA-Met"
RNA      6607
        /note="codon recognized: GAC"
RNA      6607
        /product="tRNA-Asp"

alignment_scores:
      Quality: 609.00      Length: 396
      Ratio: 2.316      Gaps: 5
      Percent Similarity: 66.414      Percent Identity: 33.586

alignment_block:
US-09-724-797-36 x AF008220 ..

Align seg 1/1 to: AF008220 from: 1 to: 220060

      8 AlaphAspProThrsAlaAspValAlaGAspProTyrProSerTy 24
      ||| ::: :|||||:|||||:|||||:|||||:|||||:|||||:
      88712 GCATGCTCAACGCACTCTCTGAGTTTGGAAAAACCCATATCTTTTAA 88761
      24 rHisTrpLeuLeuArgHisAspProValHisArgGlyAla.....H 38
      | ||| |||:|||||:|||||:|||||:|||||:|||||:|||||:
      88762 GCACACATTCGACAGCTGTTCATCCCTATCTATTAAGGAGATTTCTTAAT 88811
      38 iAsArgValTrpTyrValSerArgProAlaAspValAlaGValAlaIleuGly 54
      :: |||||:|||||:|||||:|||||:|||||:|||||:|||||:
      88812 ACCCGGCGCTGGTATGTCACAGCATGTGAAGAAACGGCTGCTATTTTGAA 88861
      55 AspGluArgPhe...AlaArgThrGlyIleAsArgProPheTrpThrAspLe 70
      ||| |||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

```

88862 GATCGGAGATTCAAAAGTCCGCGACCCCGCTGCTGAGAGCTCAACCAATA 88911
70 uValGlyProGlyLeuAlaGluIleValGlyAspIleIleLeuPheG 87
88912 TCAGGAC.....CTTTCACATGTGCAAAATCAATGATGCTGTTTC 88952
87 InaSerGluProAspHisGlyArgLeuArgGlyValValGlyProAlaPhe 103
88953 AGAACCGAGCCGATCATAGAGATGGGAGAGCGCTGGCCAGCGGAGCGCTTT 89002
104 SerProSerAlaLeuArgArgLeuGluProValIleAlaGlyThrValAs 120
89003 ACCGCGGAGAACGACAGAGATTATCCACCGCTATTCATTGCAATGTCACA 89052
120 PASLeuLeuArgProAlaLeuAlaValGlyAlaMetAspValValAspG 137
89053 TCATTTCCTTATCATAGTCAGCAAGTTAAAAAAGATGAGAGTCTTTTCGG 89102
137 LuLeuAlaTyrProLeuAlaLeuArgAlaValLeuGlyLeuGlyLeu 153
89103 ACTTGCCTTTCTTTAGCAAGTTTGTGCATAGCTAACATATATAGTGTA 89152
154 ProAlaAlaAspTrrpGlyAlaValAlaGlyArgTrrpSerArgAspValGlyAr 170
89153 CCGGAGGAGAGATGAGGAGCAATTAAAGAGGCGCTCGAGTCTCATTTCA 89202
170 gThrLeuAspArgGlyAlaSerAlaGluAspMetArgArgGlyHisAla 187
89203 AACCATTTATTTACCCGCTCAAGAAAGCATTAACAGAGCGCAATTTA 89252
187 laIleAlaGluPheAlaAspTyrValGluArgAlaLeuAlaArgArg 203
89253 TGGCTGTCCAGCATGATGCGATATTCAAGAGCTGATTCAAAAGAGAAAA 89302
204 ArgGluGlyGlyLeuAspLeuLeuAlaLeuMetLeuAspAlaHisAspAr 220
89303 CGCCACCCCTCAACAGATATGATGACGATGCTCTTGAAGGGAGAGAAAA 89352
220 gGlyLeuMetSerArgAsnGluIleValSerThrValValThrPheIleP 237
89353 GGATTAAGCTAGCGAAGAGAGAGGCGGATCTACGTCATATGCTGGGA 89402
237 heThrGlyHisGluThrValAlaSerGluValGlyAsnAlaValLeuSer 253
89403 TCGCGCGACATGAGACAAAGGTCATATTCATCAGCAATTCAGTCTTTGT 89452
254 LeuLeuAlaHisProAspGluLeuAspLeuArgArgArgProAspIle 270
89453 CTGGTCAGCATCCAGAACAGCTTTTGAACCTGAGAGAAATCCAGATCT 89502
270 uLeuAlaGlnAlaValGluGluCysLeuArgTyrAspProSerValGlns 287
89503 TATTGGTACCGGAGTGCAGGAAATGTTTACGCTATGAAGCCCGCAAGAA 89552
287 eAsnThrArgGlnLeuAspValAspValGluLeuArgGlyArgArgLeu 303
89553 TGACAGCCAGAGTTGCTGCAGAGATATGCAATCTGCGGGGTGACGATC 89602
304 ArgArgAspAspValValValValLeuAlaGlyAlaAlaAsnArgAspR 320
89603 CGTCAAGGAGAACAGTCTATCTTTGTTAGAGCGGCTAATTCAGACCC 89652
320 cArgArgTyrAspArgProAspAspPheAspIleGluArgAspProValP 337
89653 TAGCATATTTCAGAACCCCGCATGCTCTTGATATACAGAGAGTCTTATC 89702
337 roSerMetSerPheGlyAlaGlyMetArgTyrCysLeuGlySerTyrLeu 353
89703 CGCATCTTTCATTCGCGGATGCGCATGTTTGTGTTAGGGTCTCCGCTG 89752
354 AlaArgThrGlnLeuArgAlaAlaVal...AlaAlaLeuAlaArgLeuPr 369
89753 GCACGATTAGAGCGCAAAATTGCGATTAACTCTTTCGACAGCAATGCC 89802

```

```

369 oGlyLeuArgLeuGlyCysAlaSerAspAlaLeuAlaTyrGlnProArgT 386
1:::|||||  |||:|
89803 CAGCCTTATCTT.....GCGGATTTTGAATGGGGATCGCGCGCTTT 89846
386 hrMetPheArgGlyLeuAlaSerLeuProIleAlaPhe 398
|||||  |||:|
89847 TTGGAATTCGGCGCTTGAGAGAGCTGCGGTGATCTTT 89884

```

seq_name: gb_ba:SCH10

seq_documentation_block: 39524 bp DNA linear BCT 04-MAY-1999
 LOCUS SCH10
 DEFINITION Streptomyces coelicolor cosmid H10.
 ACCESSION AL049754
 VERSION AL049754.1 GI:4753846
 KEYWORDS ABC transporter; adenylosuccinate synthetase; aspartate aminotransferase; ATP-dependent protease ATP-binding subunit; cytochrome P-450 hydroxylase; fba; fructose 1,6-bisphosphate aldolase; lipase/esterase; oxidoreductase; purA; RNA polymerase sigma factor; spermidine synthase; transmembrane efflux protein; two-component regulator; two-component sensor kinase; uridine 5'-monophosphate synthase.

SOURCE

ORGANISM Streptomyces coelicolor A3(2).
 Bacteria; Firmicutes; Actinobacteria; Actinobacteridae; Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.

REFERENCE

AUTHORS 1 (bases 1 to 39524)
 Redenbach,M., Kleiser,H.M., Denaplatte,D., Eichner,A., Cullum,J., Kinashi,H. and Hopwood,D.A.
 A set of ordered cosmids and a detailed genetic and physical map for the 8 Mb Streptomyces coelicolor A3(2) chromosome
 Mol. Microbiol. 21 (1), 77-96 (1996)
 97000351

TITLE

JOURNAL 2 (bases 1 to 39524)
 MEDLINE 97000351
 REFERENCE

AUTHORS 3 (bases 1 to 39524)
 James,K.D., Parkhill,J., Barrell,B.G. and Randsdram,M.A.
 JOURNAL unpublished
 TITLE Submitted (05-MAY-1999) Streptomyces coelicolor sequencing project, Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA E-mail: barrell@sanger.ac.uk Cosmids supplied by Prof. David A. Hopwood. [3] John Innes Centre, Norwich Research Park, Colney, Norwich, Norfolk NR4 7UH, UK

JOURNAL

COMMENT

Streptomyces coelicolor sequencing at The Sanger Centre is funded by the BBSRC.
 Details of S. coelicolor sequencing at the Sanger Centre are available on the World Wide Web.
 (URL: http://www.sanger.ac.uk/Projects/S_coelicolor/)
 CDS are numbered using the following system eg SC7B7.01c (S. coelicolor), 7B7 (cosmid name), .01 (first CDS), c (complementary strand).
 The more significant matches with motifs in the PROSITE database are also included but some of these may be fortuitous.
 The length in codons is given for each CDS.
 Usually the highest scoring match found by fasta -o is given for CDS which show significant similarity to other CDS in the database.
 The position of possible ribosome binding site sequences are given where these have been used to deduce the initiation codon.
 Gene prediction is based on positional base preference in codons using a specially developed Hidden Markov Model (Krogh et al., Nucleic Acids Research, 22(22):4768-4778(1994)) and the FramePlot program of Bibb et al., Gene 30:157-66(1984) as implemented at <http://www.nih.gov/jb/>
 jun/cgl-bin/frameplot.pl. CAUTION: We may not have predicted the correct initiation codon. Where possible we choose an initiation codon (atg, gtg, ttg or (att)) which is preceded by an upstream ribosome binding site sequence (optimally 5-13bp before the initiation codon). If this cannot be identified we choose the most upstream initiation codon.
 IMPORTANT: This sequence MAY NOT be the entire insert of the

US-09-724-797-36 x SCH10/rev ..

Align seg 1/1 to reverse of: SCH10 from: 1 to: 39524

```

8 AlaPheProThrAspAlaAspValArgAspProTyrProSerTy 24
|||||
11509 GCCTTCGACCCGCGGAGACCCCGCTCGCCGACCCCTACCCGCTT 11560
24 rHsTrpLeu.....LeuArgHisAspProValHisA 35
: ||| :
11559 CGCCGAGCTCGGCGCGCGCGCTCTACTACGAGCCGAGCGACC 11510
35 rglGlyAlaHisArgValTrpTyrValSerArgPheAlaAspValArgAla 51
: : ||| ||| :
11509 AG.....TGGCTGTGTCGCCGACACACCGCGAGCTCTCCGG 11475
52 ValLeuGlyAspGluArgPheAlaArgThrGlyLeaArgArgPheTrpH 68
: ||| ||| :
11474 CTGCTGCGCGACCGCGCTCGCGCGGACTACAGACACCGTTTCACGCA 11425
68 rAspLeuValGly..... 72
: |||
11424 CGAGGACTTCGCGCGACGCCGCCGCCGCGGAGACGAGACCCCTTCACA 11375
73 .....ProGlyLeuLeuAlaGluLeuValGlyAspIleLeu 85
||| : |||
11374 CGCTCAACGACACACGCATGCTC.....GACGTG..... 11346
86 PheGlnAspGluProAspHisGlyArgLeuArgGlyValValGlyProAl 102
: : ||| ||| ||| :
11345 .....GAGCGCGCGACACACCGCTATCCGCGCTGTGTCGAAAGGC 11302
102 aPheSerProSerAlaLeuArgArgLeuGluProValAlaAlaGlyThrV 119
|||||
11301 GTTACGCGCGGACCGCTGAGCGCGCTGAAGCCCTACGTGACAGGCGCTGG 11252
119 AlaAspAspLeuArgProAlaLeuAlaArgGlyAlaMetAspValVal 135
: ||| ||| :
11251 CGGACGACCTGTGGTGGCGGTGGCGGCGCGCGGACGATCTGCTC 11202
136 AspGluLeuAlaTyrProLeuAlaLeuArgAlaValLeuGlyLeuGlu 152
: ||| ||| :
11201 ACCGACGCGGCGGAGCGCGCTGCGCGCTGATCGCCGAGATGCTGGG 11152
152 yLeuProAlaAlaAspTrpGlyAlaValGlyArgTrpSerArgAspValG 169
|||||
11151 CATCCCGGAGTCCGACCGGCGCGCTGCGCGCTGCTGGCGGAGATCT 11102
169 yArgTrpLeuAspArgGlyAlaSerAlaGluAspMetArgArgGlyHis 185
: : ||| |||
11101 GCGGAGATCTACGAGCTGAACCGTCCGAGAGACGCGCGGACGCGCTC 11052
186 AlaAlaIleAlaGluPheAlaAspTyrValGluArgAlaLeuAlaAspArg 202
: ||| :
11051 CGGCGCTCGCTCGACTTCTCGGACTACCTGCGCGCTGATCGGCGCGG 11002
202 gArgArgGluGlyGluAspLeuAlaLeuMetLeuAspAlaHisA 219
|||||
11001 CGCGAAGAGCCCGCGACGACGTATCTCGGCTGTATCCGCGCCACG 10952
219 sParg...GlyLeuMetSerArgAsnGluIleValSerThrValValThr 234
: ||| :
10951 ACGAGAGACGACCGCTCACGAGCAGAGATGATCTGACTGCGCTCGT 10902
235 PheIlePheTrpGlyHisGluThrValAlaSerGlnValGlyAsnAlaVal 251
: ||| :
10901 CTCTCGAAGCGCGGTCCAGAGGCCAGGTGAACGCGCAACACGCGGTG 10852
251 IleuSerLeuAlaHisProAspGlnLeuAspLeuArgArgArgP 268
|||||
10851 GTGGCGCTTTCGCCACCCGACGACGACTGCGCGCTTGGCGCGGAC 10802
268 rAspLeuLeuAlaGlnAlaValGluLucysLeuArgTyrAspProSer 284

```

```

10801 ACTCCCTGTCGCCGCTGCGGAGGAGCTGATGCGCTACGACACCCCG 10752
285 ValGlnSerAsnTrpArgGlnLeuAspValAlaGluLeuArgGlyArg 301
: ||| :
10751 CTCCAGCTTTTCGACGCTTGGGCTCGGACGATCGAGATCGACGGCAC 10702
301 gArgLeuArgArgAspAspValValAlaValLeuAlaGlyAlaAlaAsnA 318
||| |||
10701 GACCTTCGCGCGCGCGCGGAGCTGCGATGCGATCTGTGCGCTCGGCGAAC 10652
318 rGAspProArgArgTyrAspArgProAspAspPheAspIleGluArgAsp 334
: ||| :
10651 ACGACCGCGCGCTTTCACCGACCGGAAACGCTTCGACTACCGCGCGC 10602
335 ProValProSerMetSerPheGlyAlaGluMetArgTyrCysLeuGlySe 351
|||
10601 GACACCGCGACATCTCTTACGCGCGCGCATCCCTACTGTCATCGCGCG 10552
351 rTyLeuAlaArgThrGlnLeuArgAlaAlaValAlaAla...LeuAlaA 367
: ||| ||| :
10551 CCCGCTGGCGCCGATCGAATCTGCGCGCTCATGACGTCTGCTGAAGC 10502
367 rGLeuProGlyLeuArgGluGlyCysAlaSerAspAlaLeuAlaTyrGln 383
||| ||| ||| |||
10501 GCGCGCGCGCGCTGCGCTGCGGCGGAGACCGGAACGC.....AGG 10461
384 ProArgThrMetPheArgGlyLeuAlaSerLeuProIle 396
||| : |||
10460 CCGAATCTGTATGCGCGCGCTGAGGAGGTACCGCTG 10422
seq_name: gb_AB070947
seq_documentation_block:
LOCUS AB070947 24268 bp DNA linear BCF 11-OCT-2001
DEFINITION Streptomyces avermitilis polyketide-9 biosynthetic gene cluster.
ACCESSION AB070947
VERSION AB070947.1 GI:15824107
KEYWORDS
SOURCE
ORGANISM
Streptomyces avermitilis DNA.
Streptomyces avermitilis
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
1 (bases 1 to 24268)
REFERENCE
AUTHORS Omura,S., Ikeda,H., Ishikawa,J., Hanamoto,A., Takahashi,C.,
Shinohe,M., Takahashi,Y., Horikawa,H., Nakazawa,H., Osomoe,T.,
Kikuchi,H., Shiba,T., Sakaki,Y. and Hattori,M.
Genome sequence of an industrial microorganism Streptomyces
avermitilis: Deducing the ability of producing secondary
metabolites
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 98 (21), 12215-12220 (2001)
PUBMED 11572948
REFERENCE
AUTHORS Ikeda, H.
TITLE Direct Submission
JOURNAL Submitted (01-SEP-2001) Haruo Ikeda, Kitasato University, Dept. of
Microbial Chemistry, School of Pharmaceutical Sciences, 5-9-1
Shirokane, Minato-ku, Tokyo 108-8641, Japan
(E-mail: ikeda@mc.pharm.kitasato-u.ac.jp, Tel: +81-3-5791-6242,
Fax: +81-3-3444-6197)
FEATURES
source
location/qualifiers
1..24268
/organism="Streptomyces avermitilis"
/db_xref="taxon:33903"
/note="Actinobacteridae"
1..1875
/codon_start=1
/transl_table=11
/product="putative membrane protein"
/protein_id="BAB69274.1"
/db_xref="GI:15824108"
/translation="MSAVHFSFWIAVAVVVLMTFVVASYRQNVHAPSGGDEYVANT
NLGPRAGLTVASLLVDYLVLTVAVSINSIGLENSALPFVEHKVSCAVAVIILTLIM

```


CDS

AEAVGKEDGAVLYRPDGFVGWRTATTNNAVETSTQHILRDVLDRIILROKA"

15054 .16253
/codon_start=1
/transl_table=11
/product="cytochrome P450"
/protein_id="BAB69286.1"
/db_xref="GI:15824120"
/translation="MDPADDPYREYANILRAEPYROYLEGGILYLVREYDKILL
SDPKSKDPRNAPLDMDQAGKRPLEDRTGLTLLTTDAEPIRLRLRYSTATAR
VEGLAAQVQHTIDSLDTIVRGOAELIGDAFLPATIVICELGVPRADQVFRWT
KDFRWMTDASOADRDARDPVGRLDLEYLTRLVDRKODPADGLVDALIAADDD
RLNEAELISMSLILVGFETVNLIGNTALLRHDPQLALREPELVDALAEML
RYDGFETATRPLEPIEVAGTRIEMKHPVLLSLASNRDGAFFPAPDDPDVTRAP
AAVARGAHFICLAPLARLEGRILAFHLLRLRGLALSVPEQLRMGRSLTVAGLEA
LPVTFDA"

alignment_scores:
Quality: 563.00 Length: 406
Ratio: 2.298 Gaps: 9
Percent Similarity: 60.345 Percent Identity: 33.498

alignment_block:
us-09-724-797-36 x AB070947 ..

Align seg 1/1 to: AB070947 from: 1 to: 24268

15 AspyValrArgAspProtyrProserTyrHisTrpLeuArgHisAs 31
|||||
15057 GATTTCCGGCGCTGATCCCTATCCGAGTACCGAGTTACGTCCGGAGGA 15106
31 pProvalHis.....ArgLYAlaHisArgValTrpTyrV 43
|||||
15107 ACCGGTCCGGCAGGTCCTGGAGGGCCCGGCTTTACGGACTG...CTGG 15153
43 aLserArgpheAlaAspValArgAlaValLeuGlyAspGluArgpheAla 59
|||||
15154 TGACCCGGTAGACGAGCGTGGGAAAGCTGCTGCATCCCGGATAGC 15203
60 Arg.....ThrGlyLeuArgAr 65
|||||
15204 AAGGACCCGGCGAAGCAGCGCTGACTGGCAGAGGGGGCAAGGGAG 15253
65 pGheTrpThrAspLeuValGlyProGlyLeuLeuAlaGluIleValGly 82
|||||
15254 ACCCGTGGAGAGCCGAGACCGG.....CTCGGA 15282
82 splIleLeuPheGlnAspLeuProAspHisGlyArgLeuArgGlyVal 98
|||||
15283 CGCATCTGCTGACCAACGAGCGCCGACACACACGCGCTGCGGCTG 15332
99 ValGlyProAlaPheSerProSerAlaLeuArgTrpLeuGluProVal 115
|||||
15333 GTCTGACCGGCTTACACCGCCCGCGCTGAGGGCTGCGCGACAGGT 15382
115 ealaglyThrValAspAspLeuArgProAlaLeuAlaArgGlyAla 132
|||||
15383 GCAGACATCAACGAGCGGCTCTGCACAGATGTCGCCCGCGGAGG 15432
132 etaAspValValAspGluLeuAlaTrpProLeuAlaLeuArgAlaValLeu 148
|||||
15433 CCGAGCTCATGCGGACTTCCGCTCCGCTGCGATCAGCGTGAATCGT 15482
149 GlyLeuLeuGlyLeuProAlaAlaAspTrpGlyAlaValGlyArgTrpSe 165
|||||
15483 GAACTGCGGGAGTCCCAAGCGGACAGACGTGTCCCGCCAGTGAGAC 15532
165 rArgAspValGlyArgTrpHisAspArgGlyAlaSerAlaGluAspMet 182
|||||
15533 GAAGAACTTCGGCGGTGAGCAACACGAGCTCGCCGACGAGCGAGC... 15578
182 rGArgGlyHisAla.....AlaIleAlaGluPheAlaAspTyrVal 195
|||||
15579 ..CGGCGTACGCGCGCGGTGAGTGCAGACCTCTCGAATATCTCG 15626

196 GluArgAlaLeuAlaArgArgArgGlyGlyGlyAlaAspLeu... 211
|||||
15627 ACCCGAGTGGTGCAGACAGCGCCGTACAGACCCCGCCAGACTGGTGA 15676
212 AlaLeuMetLeuAspAlaHisAspArgGlyLeuMetSerArgAsnGlu 228
|||||
15677 CGCGGTGATCCGGCGCCGAGCAGACGACCGGCTGCAAGACGCCAGT 15726
228 LevalSerThrValValThrPheIlePheTrpGlyHisGlyThrValAla 244
|||||
15727 TGCCTCCATGATGTCCTGCTGCTGGCGGCGGCTTGCAGACAGGCTC 15776
245 SerGlnValGlyAsnAlaValLeuSerLeuLeuAlaHisProAspGlnLe 261
|||||
15777 AATCTGATCGGACAGCGCACCTGGCCCTGCTGGCCGACCCCGACCACT 15826
261 uAspLeuLeuArgArgArgProAspLeuLeuAlaGlnAlaValGluLuc 278
|||||
15827 CCCTGCTGCGGAGCGAGCGACCGAACTGTGAGTCCGCGTGGAGGA 15876
278 yLeuArgGlyArgProSerValGlnSerAsnThrArgGlnLeuAspVal 294
|||||
15877 TCGTGGGTACGAGCGGTCTTCGAGACGGCGAGCTGGCGTTCCCGCTC 15926
295 Asp...ValGluLeuArgGlyArgArgLeuArgArgAspAspValVal 310
|||||
15927 GAACCATGAGAGTGGCGGCGACGCGCATCGAAGAGGCGCACCCGCTCT 15976
310 lValLeuAlaGlyAlaAlaAsnArgAspProArgArgTyrAspArgPro 327
|||||
15977 CCGTCCCGCATGCGCCACCGGAGCGGCGAAGTTCGCGGACCGG 16026
327 spAspPheAspIleGluArgAspProValProSerMetSerPheGlyAla 343
|||||
16027 ACGACTTCAGCGTACACCGCGCGGACCCCGCCACAGTGGCTTCGGGGG 16076
344 GlyMetArgTyrCysLeuGlySerTyrLeuAlaArgThrGlnLeuArgAl 360
|||||
16077 GGTGGCGACTTCTGCTGCTGCTCCCTGGCCGCGGTGAGGCGCGGAT 16126
360 alaVal...AlaAlaLeuAlaArgLeuProGlyLeuArgLeuGlyCys 376
|||||
16127 CGGCTTCACAGCGCTGCGCGGCTGCGCGCTGCGCTGCGCTGCTGCTG 16176
376 lAserAspAlaLeuAlaTrpGlnProArgTrpMetPheArgGlyLeuAla 392
|||||
16177 CGCGGAGCAGTTGGCATGCGAGCGAGCCTCAGCGTCCGAGAGCTGGA 16226
393 SerLeuProIleAlaPhe 398
|||||
16227 GCCCTCGCGGTCACTTC 16244

seq_name: gb_ba:AE004675

seq_documentation_block:
LOCUS AE004675 10060 bp DNA linear BCT 30-AUG-2000
DEFINITION Pseudomonas aeruginosa PA01, section 236 of 529 of the complete genome.
ACCESSION AE004675 AE004091
VERSION AE004675.1 GI:9948522
KEYWORDS
SOURCE Pseudomonas aeruginosa.
ORGANISM Pseudomonas aeruginosa
Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
Pseudomonas.

REFERENCE
1 (bases 1 to 10060)
AUTHORS
Stover,C.K., Pham,X.Q., Erwin,A.L., Mizoguchi,S.D., Warren,P.,
Hickey,M.J., Brinkman,F.S., Huftagle,W.O., Kowalik,D.J., Lagrou,M.,
Garber,R.L., Goltry,L., Tolentino,E., Westbrook-Wadman,S., Yuan,Y.,
Brody,L.L., Coulter,S.N., Folger,K.R., Kas,A., Larbig,K., Lim,R.,
Smith,K., Spencer,D., Wong,G.K., Wu,Z. and Paulsen,I.T.
Complete genome sequence of Pseudomonas aeruginosa PA01, an

JOURNAL
MEDLINE
REFERENCE
AUTHORS
2 (bases 1 to 10060)
Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warrenner P., Hickey M.J., Brinkman F.S.L., Hutnagle W.O., Kowalik D.J., Lagrou M., Garber R.L., Goltzy L., Tolentino E., Westbrock-Wadman S., Yuan Y., Brody L.L., Coulter S.N., Folger K.R., Kas A., Lardi K., Lam R.M., Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T., Reitzer J., Sailer M.H., Hancock R.E.W., Lory S. and Olson M.V.
TITLE
JOURNAL
Direct Submission
Submitted (16-MAY-2000) Department of Medicine and Genetics,
University of Washington Genome Center, University of Washington,
Box 352145, Seattle, WA 98195, USA
FEATURES
Location/Qualifiers
1..10060
/organism="Pseudomonas aeruginosa"
/strain="PA01"
/db_xref="taxon:287"
70..1416
/gene="PA2472"
70..1416
/gene="PA2472"
/codon_start=1
/transl_table=11
/product="probable MFS transporter"
/protein_id="AAG05860.1"
/db_xref="GI:9948523"
/translation="MPTPDKNKTIQVAFIDQAVSAFQLRVLSICFLIIVMGEDTIA
AIGFIAPALAHGWSLPALSPILGALAGALGAFAGPLADRGKRSVLLSYLF
GMSLASAVAGSEVETLALREFTGIGLAMPNATITLSECPRRHRLAMTYMCCF
TIGSALGILARMPALMGESVILLGGPLASPLMACLPESVREFLRAPDAAR
LRRIERGLRDPDWAGRLPEADSDGTSPLRILGALMGCSLLMATPEMGLIT
VLTWMIPITLGGTGESEATITAMRGITGTCGALLIGSAMDFDHRRLSLAYG
GALFTIGASLHSPALLALCYAGVGFISGSOVQANLADFTFRKATGVSAALG
IGRIGSIVSGSGALLGLGSGSGLLALVLPALLAAVAHRGRARRSPPTL"
1429..2073
/gene="PA2473"
1429..2073
/gene="PA2473"
/codon_start=1
/transl_table=11
/product="probable glutathione S-transferase"
/protein_id="AAG05861.1"
/db_xref="GI:9948524"
/translation="MOLYSFENSSTSVFRTALAKGIDYOVVYNLRGROLPPAD
QRNPMGALPTIVDADGRPSOSLITIDVADQPPRIIPDPLHRAQALELILVAC
DIHPLANRVLYLTVOVIGIDEDRQRMVAHWAGLAAEFTLLNRHRRGAFFAGAA
GIVECCLVPOQLANARRMCDLAPYPALLEGRCLAEAFORASPEROPDYLPP"
2115..3026
/gene="PA2474"
2115..3026
/gene="PA2474"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAG05862.1"
/db_xref="GI:9948525"
/translation="KTRTRRASLLPLLLSPQALIAALPGMNLCQTSFTDGLKLPCT
LFQGFVITEADDFDANGHRRLDNRKTLTASHFAHISOHLFGAYVAAEILVY
LLADPDEFHAGSANGNLIVSPILQMPQTELRGWRQLNLAPDLPGRYDRHR
AINPGSNHSEFNPHYALIMEFAEGWEASARLOYLPGKNRDPAPPLADYDOPGAFH
ANYSVSLADAMRLGLSGYSLRQITDENVOGHDRPDSRERDAIGPALMGWGKSRLL
PLNAYHEGARRRSGSRLLIRYSQVF"
complement(3008..4342)
/gene="PA2475"
complement(3008..4342)
/gene="PA2475"
/codon_start=1
/transl_table=11
/product="probable cytochrome P450"
/protein_id="AAG05863.1"

/db_xref="GI:9948526"
/translation="MDDAFSEGSAPORRHDAORPALAPRSGDFITHTYHPDFADPEY
LRLIRSRAPVCRDOASIMWISRYADVSACLRDRRSADPARIGAGVQOGASMGCH
QOQOPLARFDNRELFENDAPRRHRLRRLPAPAGPPVAVRWERIRIYVEILDSILE
RREDFDLRDEPULTRVAEELGFPREDGQLLPGRDILAAGDLAASHGEGQINR
SAVAFSYLRQAROGMSDSSRPSGAAPILDGAAMLEGLLELVAAYAFVAA
FEETISVAGNATLLHPDQDLRCPETLANVAEELRFGAARGVCRTIEVE
IGGRIIPGGKWNLSFLAANDREFAAPRLOLORANAKOHVAFAGHVCYGLAVLA
RLQOCLRLCYRRRLASEPDLRRRSVRIERLPIYEGDAOKTCE"
complement(4415..5185)
/gene="dsbg"
/note="PA2476"
complement(4415..5185)
/gene="dsbg"
/codon_start=1
/transl_table=11
/product="thiol:disulfide interchange protein DsbG"
/protein_id="AAG05864.1"
/db_xref="GI:9948527"
/translation="MRLPRLITLGLITLLNGLATAAEELPAPIRMYOEKARILG
SFADPDLRGYAAEYQOMGALYITPGKHVLTGLFDAGKDLSPRELERIYAFILA
KEMQKESQAMTADGRADA.PRVYVLPSPDNCYCMFMEQARPWDAKCYOURLHM
GITREDESKASAILASKDPQKALHDEBQAKRSTLKPLAKTPAAVRKQLAGMEIME
SMGAATPAIFLYINAEGRMQOQOQAPDQLAEILGPR"
complement(5182..6018)
/gene="PA2477"
complement(5182..6018)
/gene="PA2477"
/codon_start=1
/transl_table=11
/product="probable thiol:disulfide interchange protein"
/protein_id="AAG05865.1"
/db_xref="GI:9948528"
/translation="MISFSIGPIVYSIOHLLFLAIGALILGGWIAARGGRAPVYL
FNLLILILVARLAFVRYRPOGRDPAQMLDLDGCFIAMPPLILAVIGALWMARR
PALRSGVAGASJGLAWLIGSLGILYERGTLPPELSTLNAGSESVQLEDFGRVLY
INLWASCPPCREMPVLOQAQENPDVVELFANQGSATVNHFLQGEMLRDLNLLF
DNGQIQOQVSVALLPTVFYTAEGRLGSHLGELSRGSLARYLEFEPAAAPATRS
SE"
complement(6018..7781)
/gene="PA2478"
complement(6018..7781)
/gene="PA2478"
/codon_start=1
/transl_table=11
/product="probable thiol:disulfide interchange protein"
/protein_id="AAG05866.1"
/db_xref="GI:9948529"
/translation="MRYLILLMLLPGLSQOAPGDDLEAPRGATQTDLPVEKAREF
TWERLDGQVQLRWQIAPGYLYQKRLRFPGDLPALQPOLPPESSHDEFGEVQYR
QSELTLPAAAGQRLGMOGCADAGCYPPQOALDGGTGAAGTSGEVAEDGL
AGSLQAGNLAMSLILPFGILLAPACSPMPILLAGIVVSGAGPRRLIAGSYV
LSMAIYAGICVVAALLGNIQWMLQOQPMILGSPALFYFLALPMGFPETLOALR
DRDGLSRKRGKSLAGAAALGALSGLLVPCMTAPLACALITYAOTGNALHGVLYF
SLGIGIMPULLLVTVGSRFLPKRGPMMNLVKGVFGFLGTAWILRLPLGLGALIG
LGGALLIVLVAALHTARGLARHAYLEGAAGCTFGMWGAMLLGAAAGADDDPPRPOV
YAAANRGAPTASHEAFLTVSQPAEILROLAAKAGQWLLDYYADMCVSCRIMEK
OYFAKPPVLAALQGVRLRLDVTADNAASSELLHRYQVPEPSSLINIGGEERRARR
LTCGVDAAGCFLAHQATRERG"
7944..8624
/gene="PA2479"
7944..8624
/gene="PA2479"
/db_xref="GI:9948530"
/codon_start=1
/transl_table=11
/product="probable two-component response regulator"
/protein_id="AAG05867.1"
/db_xref="GI:9948530"
/translation="MHVILFEDDDLASGIVAGIINAGLTVDRVASAADQALLOYAR
FDVLVYIDGLPDEGLRLRLRQOGVDPVYVTAADYATDPAVLQAGADYLLKRP
FDLRELGARLHTLQRRSAGRCVAVTEHGRISTDPSTRETYLDSRPVELSRREALLQA
LLNNGRILISGDLKDSYVGFGEVESNALNVAIHHLRRLGNAIVQTVKGLGYRLGP
ARGDGDA"

TITLE Characterization of the macrolide P-450 hydroxylase from
JOURNAL Streptomyces venezuelae which converts narbomycin to picromycin
MEDLINE Biochemistry 37 (42), 14937-14942 (1998)
98453342
PUBMED 978370
2 (bases 1 to 1470)
REFERENCE Medaniel,R. and Belbach,M.C.
AUTHORS Direct Submission
TITLE Submitted (25-AUG-1998) Kosan Biosciences, Inc., 1450 Rollins Rd.,
JOURNAL Burlingame, CA 94010, USA
FEATURES
Source Location/Qualifiers
1..1470
/organism="Streptomyces venezuelae"
/strain="ATCC15439"
/db_xref="ATCC:15439"
/db_xref="taxon:54571"
122..1372
/gene="picK"
122..1372
/gene="picK"
/function="synthesis of picromycin, methymycin, and
neomethymycin"
/codon_start=1
/transl_table=11
/product="cytochrome P450 monooxygenase"
/protein_id="AAC64105.1"
/db_xref="GI:3746884"
/translation="MRKQGGTTASPVLDGALGODEADRPYARLARAGRAHRY
RPBGEDEVLYVGYDRARVADLPFSKDMRSTPLTEAEALHNHLESPPRHTL
LRKLVARERTMARVELLRPRVEIVDGLVDAMLAPDRADLMSLAWPLPTVISEL
LGVPEDRARFVWTDVAFVDPDPAQQTANAMSGYLSRLIDSKRGDGEDLSALY
RTSDEGSRLLSEELLGNHILLVAGHETVNLINGWYALISHPDALAKADMTL
DGAVEMLRYEGPVSATYRFVPEVDLDGYIPIANGDVLVVLADAHRTPEPRPDR
FDIRDTAGHLAFGHG IHFCIGAPLARLEARLAVALLERCDLADVSPETLVYPR
PMIRKALPIMRRGRGREGRRGTG"
BASE COUNT 208 a 559 c 480 g 223 t
ORIGIN
alignment_scores:
Quality: 538.50 Length: 398
Ratio: 2.216 Gaps: 8
Percent Similarity: 61.055 Percent Identity: 34.673
alignment_block:
US-09-724-797-36 x AF087022 ..
Align seg 1/1 to: AF087022 from: 1 to: 1470
15 AspValArgAspProTyrProSerTyrHisTrpLeuLeuArgHisAs 31
188 GATTTCGGCGCGCATCCGATCCAGCTACGCCAGACTGCGTCCGCGAGG 237
31 pProValHisArg.....GlyAlaHisArgValTrpTyrValS 44
238 TCCGCGCCACCGCGGTGCGCACCCCGAGGGGAGCAGGTGCTGCTGTCG 287
44 erArgpheaAlaAspValArgAlaValLeuGlyAspGluArgpheaIaArg 60
288 TCGGCTACGACCGCGCGCGGTCTCTCGCGCATCCCGTTCAGCAAG 337
61 ThrGlyIleArgArgpHeTrpHisAspLeuValGlyProGlyLeuLeuAl 77
338 GAC.....TGGCGCAACTCCACGACTCCCTCGACCGAGGC 372
77 aglu...IleValGlyAspIleIleLeuPheGlnAspGluProAspHisG 93
373 CGAGCGCGCGCTCAACCAACATGCTGGAGTCCGACCCCGCGGACACA 422
93 IyArgLeuArgGlyValValGlyProAlaPheSerProSerAlaLeuArg 109
423 CCGGCGTCCGCAAGCTGTGTGGCCGTGAATTCAACATGCGCGGCTCGAG 472

110 ArgLeuGluProValIleAlaGlyThrValAspAspLeuLeuArgProAl 126
473 TTCTGGGCGCCCGGCTCCAGCAGATCTCGAGGGGCTGTGAGCCCAT 522
126 aleuAlaArg.....GlyAlaMetAspValValAspGluLeuAlaTyrP 141
523 GCTGGCGGGCGCCGAGCGCGCGCGCATCTGATGAGTCTCGTGGCTGGC 572
141 roLeuAlaLeuArgAlaValLeuGlyLeuLeuGlyLeuProAlaAlaAsp 157
573 CGCTCGCATACCTGTATCTCCGAACTCTCGGCTCGCGCGGCGGAGC 622
158 TrpGlyAlaValGlyArgTrpSerArgAspValGlyArgThrLeuAspR 174
623 CGGCGCGCTTCGCGCTGTGACC.....GAGCG 651
174 gGlyAlaSerAlaGluAspMetArgArgGlyHisAlaAlaIleAlaGluP 191
652 CTTCGTCTCCGCGAGATCCGCCAGGCCAGACCGCATGGCCGAGAGA 701
191 heAlaAspTyrValGluArgAlaLeuAlaArgArgArgGluGlyGly 207
702 TGAAGCGGTAATCTCCCGCTCATGACTCCAGCGCGCGGAGAGCGGC 751
208 GluAspLeuLeuAlaLeuMetLeuAspAlaHisAspArg.....GlyLe 222
752 GAGGACTCTCTACCGCGCTCGTGGCGGACCGAGGAGAGCGCTCCGC 801
222 uMetSerArgAsnGluIleValSerThrValAlaThrPheIlePheThrG 239
802 GCTGACCTCCGAGAGACTCTCGGTATGCGCACATCTGCTGTCGCGCG 851
239 IyHisGluThrValAlaSerGlnValGlyAspAlaValLeuSerLeuLeu 255
852 GGGACGAGACACGAGTCAATGTATGATCCACGAGCATGACGCTGCTC 901
256 AlaHisProAspGluLeuAspLeuLeuArgArgArgProAspLeuAl 272
902 TCGCACCCCGACAGCTGCGCGCGCGCGCGGCGGACATGACCTTTGGA 951
272 agAlaValAlaGluGluCysLeuArgTyrAspProSerValGlnSerAsnT 289
952 CGCGCGGTGAGAGATGTGTGCTGATGAGAGCGCGGAGTCCGCGGA 1001
289 hrArgGlnLeuAspValAsp...ValGluLeuArgGlyArgArgLeuArg 304
1002 CCTACCGCTTCCCGGTGAGCCCGTGCACCTGAGCGGCGACGGTCAATCCG 1051
305 ArgAspAspValValValValLeuAlaGlyAlaAlaAsnArgAspProAr 321
1052 GCGGTGACACGGTCTCTGCTGCTGCGCGCGAGCCGACCGACCCCGGA 1101
321 gArgTyrAspArgProAspAspPheAspIleGluArgAspProValProS 338
1102 GCGCTTCCGAGCCGACCGCATCTGTCGACATCCCGCGGAGACCGCGGCG 1151
338 erMetSerPheGlyAlaGlyMetArgTyrCysLeuGlySerTyrLeuAla 354
1152 ATCTGCGCTTGGCCACGCGATCCACTTGTGCATCGAGGCGCCCTTGGCC 1201
355 ArgThrGlnLeuArgAlaAlaVal...AlaAlaLeuAlaArgLeuProG 370
1202 CGGTTGGAGGCCCGATGCGCTCCGCGCGCTTCTCGAAGCGTGGCCGGA 1251
370 yLeuArgLeuGlyCysAlaSerAspAlaLeuAlaTyrGlnProArgThrM 387
1252 CCTGCGCCTGAGCGTCTCCCGCGGAGACTGCTGATGATCCGAACCGGA 1301
387 etPheArgGlyLeuAlaSerLeuProIleAlaIlePheThrProGly 401
1302 TGATCCGCGGCTCAAGGCGCTGCGATTCGCTGCGGCGGAGGA 1345
seq_name: gb_Ba:AF079139

seq. documentation block:

LOCUS AF079139 4342 bp DNA linear BCT 28-OCT-1998
DEFINITION Streptomyces venezuelae pikCD operon, complete sequence.
ACCESSION AF079139
VERSION AF079139.1 GI:3800839
KEYWORDS Streptomyces venezuelae.
SOURCE Streptomyces venezuelae.
ORGANISM Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE 1 (bases 1 to 4342)
AUTHORS Xue, Y., Zhao, L., Liu, H., and Sherman, D.H.
TITLE A gene cluster for macrolide antibiotic biosynthesis in
streptomyces venezuelae: architecture of metabolic diversity
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 95 (21), 12111-12116 (1998)
MEDLINE 98445333
REFERENCE 2 (bases 1 to 4342)
AUTHORS Xue, Y., Wilson, D., Zhao, L., Liu, H., and Sherman, D.H.
TITLE Hydroxylation of macrolactones YC-17 and narbomycin is mediated by
the pikC-encoded cytochrome P450 in Streptomyces venezuelae
JOURNAL Chem. Biol. 5 (11), 661-667 (1998)
MEDLINE 99051447
REFERENCE 3 (bases 1 to 4342)
AUTHORS Xue, Y., Wilson, D., and Sherman, D.H.
TITLE Direct Submission
JOURNAL Submitted (17-JUL-1998) Department of Microbiology, University,
420 Delaware Street SE 1060, P.O. Box 196, Minneapolis, MN 55455,
USA
COMMENT On Oct 28, 1998 this sequence version replaced gi:3777565.
FEATURES
Source location/Qualifiers
1. 4342
/organism="Streptomyces venezuelae"
/strain="ATCC15439"
/db_xref="ATCC:15439"
/db_xref="taxon:54571"
122. .4291
/gene="pikCD operon"
/note="Involved in the biosynthesis of macrolide
antibiotics methymycin, neomethymycin, and pikromycin"
122. .1372
/gene="pikC"
122. .1372
/gene="pikC"
/function="catalyzes the hydroxylation of YC-17 into
methymycin and neomethymycin and narbomycin into
pikromycin"
/note="cytochrome P450 monooxygenase"
/codon_start=1
/transl_table=11
/product="cytochrome P450 hydroxylase pikC"
/protein_id="AAC6886.1"
/db_xref="GI:3800840"

gene
1505. .4291
/gene="pikD"
1505. .4291
/gene="pikD"
/codon_start=1
/transl_table=11
/product="putative transcriptional activator pikD"
/protein_id="AAC6887.1"
/db_xref="GI:3800841"
/translation="MTLVERDGEIAHLRAVLDAAGDGTLLVSGPAGSGTELLRS
LRLAARETPWVSVALPGRDIPIGVLCQLIRAEQAGATSAVRDLIDAAHRAG
NLSPADAPLRVDETRHLHDWLISVSRPTPLVAVDILTHADTASLRFLLYCAHHDD

GGIGFVMTERRASQAGYRVFAELLRPHCRNMWLSGLPPSGVROLLAHYGPAAER
RAPAYHATGNNPLLRALRODROASTTTGAAAGDEPVAFAVDLCILRSAG
TLTETAPLAVEOSDPLIVERIGTTAAAEHRIHLEAIGLDEGTLGQPAIREAA
LODI/PASERIEELHRAAEQOLHROGADBDYVLRHILVYASGRMNPHTTALPDLSSGL
DDRIDDAPRIIEFVRSSTIDTQIARLAPRILVYASGRMNPHTTALPDLSSGL
PPSHVVALIRCLVWYGRLEPAADALSRLRPSNDLLESLTRMVAALCPPLLSL
PATPEPERGPVPRVLRPTALQAGVFORPDMASVAQEOILOGCRLEETFEAL
ETALLVHADRLRALFMSDALIAEVEERRSIGMEVFAATRAMIARICGDLPTARE
RAELALSHAPEWSGLAVGMDLSALLACTEAGEYEOAEVLRQVDAMFDSRHGME
YMARGRVWLTAGRLHVALGEPMLCGEILGSMTDPSITVPMPTSAEYVLRIGNOK
ARALAEQOLALVPRGRSRTGRTITRVIAAVDQOARLAEAVMDLHDSGDLREAR
ALAGMSRHQOAGQNVNRYARMTARLAGMAMACGAYPLAEITVPRGGRKAYSTLE
LPGRPDVGLTSEAEERRVALAARGLTRQIARLRCLVASTVEQHLTRYVRKLVNTRRA
DLPLSLADKXSVTA"

BASE COUNT 586 a 1699 c 1485 g 572 t
ORIGIN

alignment_scores:
Quality: 538.50 Length: 398
Ratio: 2.216 Gaps: 8
Percent Similarity: 61.055 Percent Identity: 34.673

alignment_block:
US-09-724-797-36 x AF079139 ..
Align seg 1/1 to: AF079139 from: 1 to: 4342

15 ASPVALARGARGSPROTYRPROSERTYRHISTRLPLEUVALRHIS 31
188 GATTTCGGCGCGCGATCGATCCGACGTACGACGAGCTGTCGCGAGGG 237
31 PPROVALHISARG.....GLVALAHISARGVALTTRTYVALS 44
238 TCCGGCGCCACGGGTCGCGACCCCGAGGGGAGAGGTGCGTCGTCG 287
44 ERARGPHEALASPVARGALVALLEUGLYASPLUARGPHEALARG 60
288 TCCGCTACGACCGCGCGCGCTCGCCGATCCGCGACCAAG 337
61 ThrGlyLeuArgPheTrpThrAspLeuValGlyProGlyLeuLeuAl 77
338 GAC.....TGGCGCAACTCCAGCACTCCCTGACCGAGGC 372
77 AGU...LLEVALGLYASPLLELEUPHEGNAISPGIUPROASPHISG 93
373 CGAGCGCGCGCTCACCAACAATGCTGGAGTCCGACCGCGCGGACACA 422
93 LYARGLEUARGLYVALGlyProAlaPheSerProSerAlaLeuArg 109
423 CCGCGCTCGCAAGCTGTCGCGCGTGCAGTTCACCATGCGCGCGGTCGAG 472
110 ArgLeuGluProValIleAlaGlyThrValAspLeuLeuArgProAl 126
473 TTGCTGGGCGCGCGGTCACGAGATCGTCGACGGCTGTGGACGCCAT 522
126 ALEUALARG.....GLVALAMELASPVARGALVALASPGIULEUALTYR 141
523 GCTGGCGCGCGCGCGCGCGCGCGATCGAGGAGTCCCTGCGCTGCG 572
141 ToleuAlaLeuArgAlaValLeuGlyLeuLeuGlyLeuProAlaAlaAsp 157
573 CGCTGCCGATCACCGTATCTCCGAATCTCTCCGCGTGGCGCGGAC 622
158 TrpGlyValAlaGlyArgTrpSerArgAspValGlyArgThrLeuAsp 174
623 CGCGCGCGCTTCCGCGCTGGACC.....GAGCG 651
174 gGlyAlaSerAlaGluAspMetArgArgGlyHisAlaAlaIleAlaGluP 191
652 CTTCGCTTCCGCGAGATCCGCGCGAGCGCGACCGCATGCGCGAG 701
191 heAlaAspTyValGluAlaGlyAlaLeuAlaArgArgArgGluGly 207


```

      702   TGAAGCGCATATCTCTCCGCGCTCATGCACTCCAGCGCGGAGGACGCG 751
      208   GLAASPLeuAlaLeuMetLeuAspAlaHisAspArg.....GlyLe 222
      752   GAGAGCTGCTCAGCGCGCTCGTGGCGACGACGAGGAGCGCTCCG 801
      222   uMetSerArgAsnGluLeuAlaSerThrValThrPheIlePheThrG 239
      802   GCTAGACTCCGAGAGACTGCTCGTATGCGCACATCTGCTGTCGCGG 851
      239   LysHisGluThrValAlaSerGlnValGlyAsnAlaValLeuSerLeu 255
      852   GGCACGAGACCGACGTCATCTGATCGCACGCGCATGACGCTGCTC 901
      256   AlaHisProAspGlnLeuAspLeuLeuArgArgProAspLeuAla 272
      902   TCGGACCCGACGAGCTGGCGCGCTGGGCGCGACATGACGCTTTGA 951
      272   aGlnAlaValGluGluCysLeuArgTyrAspProSerValGlnSerAsn 289
      952   CGGCGCGCTGAGGAGATGTTGCGTACGAGGCGCGGATCCGCGA 1001
      289   hArgGlnLeuAspValAsp...ValGluLeuArgGlyArgLeuArg 304
      1002   CTTACCGCTCCGCGTCCGACCGCGTGCACCTGGACGCGACGTCATCCG 1051
      305   ArgAspAspValValValValLeuAlaGlyAlaAlaAsnArgAspPro 321
      1052   GCGCGTGCACGCGTCTGCTGCTGCTGCGCGACGCGCACCGACCCGA 1101
      321   gArgTyrAspArgProAspAspPheAspIleGluArgAspProValPro 338
      1102   GCGCTTCCGCGACCGCGACCGCTTCCGACATCCGCGGACACCGCGCC 1151
      338   erMetSerPheGlyAlaGlyMetArgTyrCysLeuGlySerTyrLeuAla 354
      1152   ATCTGCGCTTCCGCGACGCGCATCTGCTGCTGCTGCGCGCGCGCTGCG 1201
      355   ArgThrGlnLeuArgAlaAlaVal...AlaAlaLeuAlaArgLeuPro 370
      1202   CGGTTGGAGCGCGGATCGCGCGCTTCTGAAACGCGCGCGGA 1251
      370   yLeuArgLeuGlyCysAlaSerAspAlaLeuAlaTyrGlnProArgThr 387
      1252   CCTGCGCTTGGAGCTGCTCCCGCGGACACTCGGTGTATCCGAACCGGA 1301
      387   etPheArgGlyLeuAlaSerLeuProIleAlaPheThrProGly 401
      1302   TGATTCGGGGCTCAAGGCGCTCCGATCCGCTGCGCGGAGGA 1345
      seq_name: gb_ba:SERCP450A

seq documentation block:
LOCUS      SERCP450A                      3078 bp    DNA       linear    BCT-26-Apr-1993
DEFINITION Saccharopolyspora erythraea ORF 1 gene, partial cds; cytochrome
P-450 gene, complete cds; ORF 2 gene, partial cds.
ACCESSION   M83110.1 GI:152682
VERSION     M83110.1
KEYWORDS    cytochrome P450.
SOURCE      Saccharopolyspora erythraea (library: NRRL 2338) DNA.
ORGANISM    Saccharopolyspora erythraea
             Bacteria; Firmicutes; Actinobacteria; Actinobacteriales;
             Actinomycetales; Pseudonocardiales; Pseudonocardiales;
             Saccharopolyspora.
REFERENCE   1 (bases 1 to 3078)
AUTHORS    Andersen,J.F. and Hutchinson,C.R.
TITLE      Characterization of Saccharopolyspora erythraea cytochrome P-450
           genes and enzymes, including 6-deoxyerythronolide B hydroxylase
JOURNAL    J. Bacteriol. 174, 725-735 (1992)
MEDLINE    92121109
FEATURES
             source          location/Qualifiers
             source          1..3078
```

```

CDS
      /organism="Saccharopolyspora erythraea"
      /db_xref="taxon:1836"
      /tissue_1ib="NRRL 2338"
      <1..859
      /note="ORF homologous to Escherichia coli and Salmonella
      typhimurium dnae gene; putative"
      /codon_start=2
      /transl_table=1
      /protein_id="AA26482.1"
      /db_xref="GI:152683"
      /translation="AGGGRGRRRAQAPARAPVRPVRRRRRGGVVAAGAPAVHAG
      PAAQADAILRAGDAPRLRGPSPGRRRAARAPGRRRRARRRRRRGGGRRGDR
      DDLDAPPHODERPPVDDRPDPPRRRGAVYLPYLVRRLRPAAGHDDRRGRR
      ARGHHQRLRLRGRRGHLGGRRDGNPNRVDGAGGPPVARRAQAHLGALGNG
      PGAREAGACAGHPRAVORLRLVDRERRAAGRAAGRVLRDGAIVSRGAASPLRA
      PRPLRLRGS"
      949..2166
      /codon_start=1
      /transl_table=1
      /product="cytochrome P-450"
      /protein_id="AA26483.1"
      /db_xref="GI:152684"
      /translation="MTTGEVVDLLAFDDDAFADODRHNRYARMRPEVQIRTVNGDAM
      LITREYVKQALDPRIAKDFGRTOQIIEKRLDAERPEGSPDLGPHMLNDDPDT
      RLRLVYKAFARAREGLRPIEDITDLDRLAGSEVDLIDEPAPLPIYISLM
      GVERSDRDPFRSMTNVLVDGSOPEAOASVAMVEYLTEIARKRPEDDLLTLE
      AVEGDRLSBESEILAMVFLLLVAGHETTVLTGNCVLSLIGNDOLALRNDPSLPG
      AIEELRTESFVANGTRHTAEAREDDVYIPGSELVWALGAANDGRPEPDRFD
      ITRRTTGLVAFRHGIFHCVGAALRLAEOLAIVGRLLERFPDLRMAASPDLLKRRSVL
      MRGELKLPVRPGA"
      complement(2365..>3078)
      /note="ORF 2"
      /codon_start=1
      /transl_table=1
      /protein_id="AA26484.1"
      /db_xref="GI:152685"
      /translation="LIAVANGSTLEVLSQLVGRHSADVAGMAALPLAAGSVGSA
      LGAAPARIGYRACIVGFAVTAGFVLAALPDSGPHIAGLLICGFGFAPPA
      ASTALGAVPADRAGMGLNDTHOQGLIAGVAFVAGLSTYRAFLPTGVPHDAST
      SLAATLSPADRETSALADAARLAFTQAGSATMTAGIACALAGAAVAMLSLSGRASA
      VALTRVSTATARNLRAGST"
      418 t

BASE COUNT      463 a      1060 c      1137 g      418 t
ORIGIN

alignment_scores:
      Quality:      531.00      Length:      419
      Ratio:      2.019      Gaps:      13
      Percent Similarity:      62.768      Percent Identity:      34.129

alignment_block:
US-09-724-797-36 x SERCP450A ..
Align seg 1/1 to: SERCP450A from: 1 to: 3078

3 ValAspAlaValThrAlaPheAspProThrAspAlaAspValArgArgAs 19
|||||
64 GTGCCGCACTTACTGCTTTCAGACGCGCTTCCGCGACCGCGACCA 1013
19 pProTyrProSerTyrHisThrPheLeuArgHisAspProValHisArg. 35
|||||
1014 CCGGTACGCGCGG.....ATGCGCGAGGAGCGCGGTGACGCGCA 1051
36 .....GlyAlaHisArgValThrTyrValSerArgPheAlaAsp 48
|||||
1052 TCCGCAAGGTGAACGGCGCTGCAGCGGTGATCACCGCTACGAGGAC 1101
49 ValArgAlaValLeuGlyAspGluArg.....PheAlaArgThr 61
|||||
1102 GTCAAGACGCGCTGTCGACCGACAGATCGCCAGAGACTTCGGGCGCAC 1151
61 rGly.....IleArgArgPheThrPheAspLeuVal...GlyProGlyL 75
|||||
```


1152	CCAGCGAGTCAATCGAGAGAGCGCTGGCGGACGCCGAGCGGACGGCGAGCGGCGG	1201
75	eulneuAaagluileValaJlyaspillelleuepneghinaspgluProasp	91
1202	TTCTCGCGCCGACCTC...GGCCCGGACATGCTCAACACCGGACCGCGGAGC	1248
92	HisGlyArgLeuArgGlyValValaJlyProIapheserProserAlaIe	108
1249	CACACCCGGTTCGCGAACCTGGGTGTCAAGCGCTTCACCGCGCGCGGT	1298
108	uArgArgLeuGluProValIleAlaGlyThrValAspAspLeuLeuArgp	125
1299	CAGAGGCGCTCGCGCGCGGATCGACACATACGACGACGACCTGGTCGAC	1248
125	roAlaLeuAlaArgGlyAlaMetAspValValAspGluLeuAlaTyrPro	141
1349	GCCTGGCGGGCGCGCTCCGAGAGGTGATCGACGAGTTCCGCTCCCG	1398
142	LeuAlaLeuArgAlaValLeuGlyLeuLeuGlyLeuProAlaAlaAspTr	158
1399	CTCCCATACGCTGATCTCCGAGAGTGTATGGGCGGTGAGAGACTGGCGG	1448
158	pgAlaValaJlyArgTyrSerTyrAspValaJlyArgTyrThrLeuAspArg	175
1449	CGACGACCTTCGGTGTGTGAC...AACGCTGTGTGACG	1486
175	lyAlaSerAlaGluAspMetArgArgGlyHisAlaAlaIleAlaGluPhe	191
1487	GGTCGAGCGCGGCGGCGCGGCGCGCGCTGGCGGCGATGTGTGAG...1533	
192	AlaAspTyrValaGluArgAlaLeuAlaIaArgTyrArgTyrGluGlyG1	208
1534TACCTGACCGAGCTGATCGCCAGAGAGCCGACCGAGCCGGTGA	1577
208	uAspLeuLeuAlaLeuMetLeuAspAla...HisAspArgGlyLeuMet	224
1578	CGACCTGCTGACCGCGCTGTGGAGCGGTGAGAGCGGACGACCGGCTCA	1627
224	eArgArgGluIleValSerThrValValThrPheIlePheThrGlyHis	240
1628	CGCGAGCGCGAGCTGATCGCCATGAGTTCCTGCTGCTGCTGGCGGCGAC	1677
241	GluThrValAlaSerGluAlaGlyAsnAlaValaLeuSerLeuAlaAlaI	257
1578	GAGCGACGGTCAACTGATCGGCAACTGCGGTGCTGCTGCTGGCGGA	1727
257	sProAspGluLeuAspLeuLeuAlaArgTyrArgProAspLeuAlaAla	274
1728	CCCCGACACACTGCGCGCGCTGGCGACAGACCGCGTGGTGGCGCGGCG	1777
274	IaValGluGluCysLeuAlaArgTyrAspProSerValaGluSerAsnThrArg	290
1778	CGATTCGAGGAGACCTCTGCGCTACGAGAGCCGCGTGGCGACAGCGTTC	1827
291	GluLeuAspAlaAspValaGluLeuAlaArgGlyArgGlyLeuArgArgAsp	307
1828	CGGACACCGCCGAGGCG.....GTCCGGTTCGCGGA	1859
307	pValIaVala.....ValleuAlaGlyAlaAlaAla	317
1860	CGTGTGATTCGCCGAGGAGAGCTGCTGTGGGTGCGCTCGGGGCGGCGGA	1909
317	snArgAspProArgTyrTyrAspArgProAspAspPheAspIleGluArg	333
1910	ACCGGAGCGGAGACGCTTGAGAGACCCGACCGCGTGCACATTCACCGCG	1959
334	AspProValProserMetSerPheGlyAlaGlyMetArgTyrCysLeuG1	350
1960	GAGACACACCGGCGACGTCGCTTCGCGCCACGCGCATCTTCGCTGGG	2009
350	ySerTyrLeuAlaArgThrGluLeuArgAlaAlaValaAla...AlaLeuA	366
2010	CGCGCGCCCTCCCGGTGAGAGCCAGATGTCGCGCTGGCGGACGCTCTGG	2055

[illegible]

1
134093 AAGTGCACACCGCAGCCGCCCGCTGGTGTGAAGACGCTTCCTT 134044
281 yfAspProSerValGlnSerAsnThrArgGlnLeuAspValAspValGln 297
134043 ATGATCCAGCAATCATCTAATCGTGGGGTGTGCTGATAGACATGACG 133994
298 LeuArgGlyArgArgLeuArgArgAspAspValValValLeuAlaG1 314
133993 ATCGGTCACAGACACTTTCACCGAGGCTGTACCAATGGTTGTGTCTTCG 133944
133943 CGCCGCCAATGCGCATTCACCGGCTTACACGCCCATGAATTGGAC 133894
331 LeuIAspAspProValProSerMetSerPheGlyValGlyMetArgTyr 347
133893 CCGACCGCTCCAGCTCACGACATTTGGCATTTGGCGCTGCACATTTT 133844
348 CysLeuGlySerTyrLeuAlaArgThrGlnLeuArgAlaValAlaAla 364
133843 TGCCTAGAGAGCGGCATTCGCTCGTTGGAAGACGATGCATCATATATGCG 133794
364 ileu...AlaArgLeuProGlyLeuArgLeuGlyCysAlaSerAlaAla 380
133793 GATCAGTGGCGGCTTCCACAGSTACGCTG.....GCCGTGAT 133753
380 euAlaTyrGlnProArgThrMetPheArgGlyLeuAlaSerLeuProIle 396
133752 TGGTCTATACCAATATGTGCAATCGCGGATGTGCGGACCTGCCGTA 133703
seq_name: gb_ba:SCE41
seq_documentation_block:
LOCUS SCE41 36028 bp DNA linear BCT 02-NOV-2000
DEFINITION Streptomyces coelicolor cosmid E41.
ACCESSION AL442120
VERSION AL442120.1 GI:10241774
KEYWORDS ABC transport system ATP-binding protein; ABC transport system integral membrane protein; cytochrome P450 hydroxylase; eno; enolase; hydrolase; integral membrane protein; lipoprotein; nucleotidyltransferase; oxidoreductase; pkat; secreted protein; sensor kinase; transcriptional-repair coupling factor.
SOURCE Streptomyces coelicolor.
ORGANISM Streptomyces coelicolor
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae; Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE 1 (bases 1 to 36028) Redenbach,M., Kiser,H.M., Denapate,D., Eichner,A., Cullum,J., Kinashi,H. and Hopwood,D.A. A set of ordered cosmids and a detailed genetic and physical map for the 8 Mb Streptomyces coelicolor A3(2) chromosome Mol. Microbiol. 21 (1), 77-96 (1996)
TITLE 2 (bases 1 to 36028) 97000351
JOURNAL 3 (bases 1 to 36028) Unpublished
MEDLINE Saunders,D.C. and Harris,D.
REFERENCE 3 (bases 1 to 36028) Direct Submission
AUTHORS Cerdano,A.M., Parkhill,J., Barrrell,B.G. and Rajandream,M.A.
JOURNAL Submitted (19-SEP-2000) Streptomyces coelicolor sequencing project, Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge
TITLE David I.A. Hopwood, [3] John Innes Centre, Norwich Research Park, Colney, Norwich, Norfolk NR4 7UH, UK
COMMENT Notes:
Streptomyces coelicolor sequencing at The Sanger Centre is funded by the BNSC and Beowulf Genomics
Details of S. coelicolor sequencing at the Sanger Centre are available on the World Wide Web.
(URL: <http://www.sanger.ac.uk/Projects/S.coelicolor/>) CDS are numbered using the following system eg SC78V.01c. SC (S. coelicolor), 787 (cosmid name), .01 (first CDS), c (complementary

strand). The more significant matches with motifs in the PROSITE database are also included but some of these may be fortuitous. The length in codons is given for each CDS. Usually the highest scoring match found by fasta -o is given for CDS which show significant similarity to other CDS in the database. The position of possible ribosome binding site sequences are given where these have been used to deduce the initiation codon. Gene prediction is based on positional base preference in codons using a specially developed Hidden Markov Model (Krogh et al., Nucleic Acids Research, 22(22):4768-4778(1994)) and the FramePlot program of Bibb et al., Gene 30:157-66(1984) as implemented at <http://www.nih.gov/jp/jun/cgi-bin/frameplot.pl>. CAUTION: We may not have predicted the correct initiation codon. Where possible we choose an initiation codon (atg, gtg, ttg or att) which is preceded by an upstream ribosome binding site sequence (optimally 5-13bp before the initiation codon). If this cannot be identified we choose the most upstream initiation codon. IMPORTANT: This sequence MAY NOT be the entire insert of the sequenced clone. It may be shorter because we only sequence overlapping sections once, or longer, because we arrange for a small overlap between neighbouring submissions. Cosmid E41.

FEATURES

SOURCE

1..36028
Location/Qualifiers

/organism="Streptomyces coelicolor"
/db_xref="taxon:1902"

gene

/gene="SCE41.01c"
complement(1..951)
/gene="SCE41.01c"
complement(<1..951)

CDS

/note="SCE41.01c, probable oxidoreductase (fragment), len: >317 aa; similar to SW:DHNA_ECOLI (EMBL:V00306) Escherichia coli MADH dehydrogenase (EC 1.6.99.3) Ndb, 433 aa; fasta scores: opt: 361 z-score: 417.8 E(): 9.1e-16; 27.6% identity in 330 aa overlap and to TR:CA892372 (EMBL:AL356612) Streptomyces coelicolor putative NADH dehydrogenase SCD72A.05, 442 aa; fasta scores: opt: 1185 z-score: 1267.3 E(): 0; 57.0% identity in 321 aa overlap. Contains Pfam match to entry PF00070 pyr_redox, Pyridine nucleotide-disulphide oxidoreductase"
/codon_start=1
/transl_table=11
/product="putative oxidoreductase (fragment)"
/protein_id="CAC09533.1"
/db_xref="GI:10241775"
/translation="MRYGEATVTVDPDSYMTYQPLPEAAGSISPRHVVPILRPVL PKAEVLGRVTTIDIDRKVAVPILVGEAYELPEQVLIAMGAVSTPPFPGIAEGIT GMKIESTIGLRNHVLEOLDKADSTDEETRRKALFFVGGFAGAEITGEVEDMAR DKAHYNNNSREDMRFTLDADKTIILEVGPKGIGYKHEHLEGRVEVLTSMDSVCV DGHVVLKNGLEVDNSNTIVTAGVKNPMPALAREGLPLGPRGHVDTQATLQVGTDYIWA AGDNAQVPDLVGRKAGNENAMPNQHALLRQAKVIGDNVI"
1..101
/note="nominal overlap with Streptomyces coelicolor cosmid SCE25"

misc_feature

SOURCE

1..36028
/organism="Streptomyces coelicolor A3(2)"
/strain="A3(2)"
/db_xref="taxon:100226"
/clone="cosmid E41"
complement(112..798)
/gene="SCE41.01c"
/note="Pfam match to entry PF00070 pyr_redox, Pyridine nucleotide-disulphide oxidoreductase, score 76.80, E-value 6.7e-22"

misc_feature

gene
complement(1477..2418)
/gene="SCE41.02c"
complement(1477..2418)
/gene="SCE41.02c"
complement(1477..2418)

CDS

/note="SCE41.02c, possible hydrolase, len: 313 aa; similar to N-terminal region of SW:GPPA_ECOLI (EMBL:M87049) Escherichia coli guanosine-5'-triphosphate,3'-diphosphate pyrophosphatase (EC 3.6.1.40) GppA, 494 aa; fasta scores:

opt: 311 z-score: 357.2 E(): 2.2e-12; 28.4% identity in 310 aa overlap"
/codon_start=1
/transl_table=11
/product="putative hydrolase"
/protein_id="CAC09534.1"
/db_xref="GI:10241776"
/translation="MTRVAVDCGTNSIRLVLADDPATGELTDDRRMTVRLGGV DRRGLAPALEETFAACREAVENVAHAEELRVASASADANRPDVFGLDITL GVEPEVTSDDAEFEFTGATKEITGRALDPKLVLDVGGSTFVEYGEEDHVARAS VDVGCCRMTERHRLVRGAVTDPPEOVAVAMRADLEALDLAGRTVPVGEATYIGLA GSATYSAIAGLEPEYDSAAIHHSRVSRRVETIDMLASTHAERAAVASHPERVD VVAGSVLVAIMERGAEVYSEHDITDGIAMSTA"
complement(2415..2975)
/gene="SCE41.03c"
complement(2415..2975)
/gene="SCE41.03c"
complement(2415..2975)
/note="SCE41.03c, conserved hypothetical protein, len: 186 aa; similar to TR:P96375 (EMBL:292539) Mycobacterium tuberculosis hypothetical 16.6 kDa protein MTCY10G2.24c, 135 aa; fasta scores: opt: 632 z-score: 730.7 E(): 0; 67.6% identity in 136 aa overlap"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="CAC09535.1"
/db_xref="GI:10241777"
/translation="MQTPPTPTPTPEPDADVAFAKQOLGRPPGRLAIRHRCGQP DVEETAPRLPDGTPPTPLLYLCPKPAASAIQTLNANGVKKETERLADPELAAAYRA AHEDYTRRDEIEELTGFPASGMPDRVKCLHVVAHSLAAGPVPNPLGDIAIMLPE WMRKGCVPTPTDDETDEGTGEDAQ"
complement(2425..2434)
/gene="SCE41.03c"
complement(3061..3585)
/gene="SCE41.04c"
complement(3061..3585)
/note="SCE41.04c, hypothetical protein, len: 174 aa; similar to TR:P96376 (EMBL:292539) Mycobacterium tuberculosis hypothetical 24.6 kDa protein MTC10G2.25c, 228 aa; fasta scores: opt: 273 z-score: 332.6 E(): 5.1e-11; 35.6% identity in 149 aa overlap. Contains possible coiled-coil region at approx residues 87..106"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="CAC09536.1"
/db_xref="GI:10241778"
/translation="MCGDVRACTGGGDMAYAKDBDRESTATRIRIIGOTARVYRSO TKQARRSRRLGRALLAMVLCISLVVALAYPIROYVAGRAEIDALOREORETRQRYVD LRDKARMDADVAEQVRLMLHYMGEITGFVYVDPPEAAGQTARAGAADRPYQNV MDGVVKAANAVARQ"
complement(3639..4919)
/gene="eno"
/note="SCE41.05c"
complement(3639..4919)
/gene="eno"
complement(3639..4919)
/note="SCE41.05c, eno, enolase, len: 426 aa; similar to SW:ENO_ECOLI (EMBL:X82400) Escherichia coli enolase (EC 4.2.1.11) Eno, 431 aa; fasta scores: opt: 1603 z-score: 1807.6 E(): 0; 60.6% identity in 419 aa overlap. Contains Pfam match to entry PF00113 enolase, Eno1-ase and match to Prosite entry PS00164 Enolase signature"

CDS

gene

/transl_table=11
/product="enolase"
/protein_id="CAC09537.1"
/db_xref="GI:10241779"
/translation="MPSIDVYVAREIIDSNGNPVEYVGLIDGSGTGRAVPSGASNG AFEAIELRDGDPSTYLGKGVKAVLAVIYDGPETVGVDTGRTQRLDQANFDDAIDN KGSIGANVAIIGVSAVHAASASADIDLEFRILGSPNALLPVPMMNTLNGSADSNV DIOGFMTAPIGAEFSFALRMGAENVHTLKKVLKNGLATGADDEGGFANLGSNREA LDIILEAIKRENGYVKGBOIALALDVAASEFYKDGSAFEGKNRSAAEMTEYVELVEA


```

predicted by Codon_usage
predicted by Homology
predicted by Framed"
/codon_start=1
/codon_start=1
/evidence=not_experimental
/product="PROBABLE ENOYL-COA HYDRATASE PROTEIN"
/protein_id="CAD17799.1"
/db_xref="GI:17431119"
/translation="MIALDWLHDCVTALLTSLRPPNAAFTPEGLQCAVETLADP
RYRALVITIGDSPKFEASADLTFASGDHVDVAREAAARGAAFEALQMARVITAIIN
GVMGSGLEACLDIRIAERHQLAVPEPVALGCCGTGTLPEMLWSEEMARKMLI
TGERIDAQTFALRIGLVEEVDDGAAREALAMARVYTLSPRAVFSKALLHQARHG
PTALAAVERERFDLGHDPDREGVNAFLKRPPIWHTAAPEALP"
complement(4302..5198)
/gene="mmsB"
/note="RSP0649; RS05573"
complement(4302..5198)
/gene="mmsB"
/EC_number="1.1.1.31"
/function="small molecule metabolism; amino acid
biosynthesis; isoleucine/valine"
/note="Product confidence : probable
gene name confidence : putative
predicted by Homology
predicted by Codon_usage
predicted by Framed"
/codon_start=1
/translation="
/evidence=not_experimental
/product="PROBABLE 3-HYDROXYISOBUTYRATE DEHYDROGENASE
OXIDOREDUCTASE PROTEIN"
/protein_id="CAD17800.1"
/db_xref="GI:17431120"
/translation="MKIAFGLGNGMAPRNILKAGHALSYFDLDAKAVOSLVDAGA
SAAPSRKAATGADCAITMLPAAAHVSLTDDGLIAGIKVPIIDSSIDPDSAK
ALAAATKAGNPEVDAPVSGSGTGAAGAGLITFVWGGAIVPEQVPILSAMGRVLYC
GEAGGCGAKIKCNLUYIGITMAGYAAEMSLGERLIGIDRVYLAGIINTSTGCMSSDTY
NFFPEVITAPSRIGYGGFTDMLKDLGLGADAKSVROPYVIGALADOLYQASS
OCAGLDESAIVRILRYRPAGAR"
complement(5220..6767)

```

gene

alignment_scores: Quality: 509.50 Length: 387

 Ratio: 1.952 Gaps: 7

Percent Similarity: 67.442 Percent Identity: 33.850

alignment_block:

US-09-724-797-36 x AL646080 ..

Align seg 1/1 to: AL646080 from: 1 to: 190050

19 AsProctyYrProSerTyRHisTrpleuLeuAgnHisAspProvalHisAr 35

 :::||||| |||::: ||| ::: |||

77427 AATCCCATCCCTCGTACGAGCGCTTCCTCGAGGCGCCCTTCGTCGCG 77476

35 gGlyAlaHisArgValTrpTyValSerArgPheAlaAspValArgAlaY 52

 ||| ::: ::: ::: |||

77477 CATGGCGCGCAATCCGTTGATGCGGGCGACATACAGCATCTGCAGCGCC 77526

52 aLeuGlyAspGluArgPheAlaArgThr.....GlyIleArgArg 65

 ::|::| ::: ::: ::: |||

77527 TGCTTGACACACCGGACATGGCAAGACTCATGAGAGCATCCGCTG 77576

66 PheTrpPhrAspLeuValGlyProGlyLeu.....LeuAlaGluIleValGl 81

 ::: ||| |||||::: ::: ||| ::: |||

77577 CGATACGGCGC...GAAGGCCCAACATGCGCGTGTTCAGGGGGTTCAG 77623

81 YAspIleIleLeuPheGlnAspGluProAspHisGlyArgLeuArgGlyY 98

 Y:::||||::: ::: ||| ||| |||||::|::|::|

77624 CCGAGTGTCTCTAATCTCAATCCGCCCATGACACAGCGCGCTGGCGGCC 77673

98 aValGlyProAlaPheSerProSerAlaLeuArgArgLeuGluProVal 114

[illegible]


```
seq_name: gb_pat:ARI44763
seq_documentation_block:
LOCUS ARI44763 6085 bp DNA linear PAT 08-AUG-2001
DEFINITION Sequence 4 from patent US 6210935.
ACCESSION ARI44763
VERSION ARI44763.1 GI:15106630
KEYWORDS
SOURCE
ORGANISM
REFERENCE
1 (bases 1 to 6085)
AUTHORS Schupp,T., Engel,N., Bietenhader,J., Toupet,C. and Pospiech,A.
TITLE Steuoridin biosynthesis gene clusters
JOURNAL Patent: US 6210935-A 4 03-Apr-2001;
FEATURES
source location/Qualifiers
1..6085
/organism="unknown"
BASE COUNT 862 a 2374 c 1992 g 833 t 4 others
ORIGIN

alignment_scores:
Quality: 508.50 Length: 407
Ratio: 2.101 Gaps: 11
Percent Similarity: 59.459 Percent Identity: 36.364

alignment_block:
US-09-724-797-36 x ARI44763 ..

Align seq 1/1 to: ARI44763 from: 1 to: 6085

20 ProTyrProSerTyrHisTrpLeuAlaArgHisAspProValHisArg.. 35
|||||
441 CCTACCCGGGTCTACCGGCTACCGGAGCGCCCGGCTCCATCGCAC 490
36 ..... GlyAlaHisArgValTrpTyrValSerArgPheAla 48
|||||
491 GGGCGGCGGCGCGGAGAGCCTGACACCTTACGTTCACCTACGACG 540
48 sPValArgAlaValLeuGlyAspGluArgPheAlaArgThrGlyLeuArg 64
|||||
541 ACGTGGTCCGGTCTGTCACACGCGGCGTTCGCGCCAGCAGCC..... 584
65 ArgPheTrpHisAspLeuValGlyProGly..... 74
|||||
585 CGCGTGGCTTCGCGGACACCGCGCCCGACACCGCGCGCTCCCGATCCC 634
75 ..... LeuLeuAlaGluIleValGlyAspIleIleLeuPheG 87
|||||
635 GGGCGAGCAGCGCGCCTCGCGGACCTGTCGAGAACTGGCTGCTTCC 684
87 InAspGluProAspHisGlyArgLeuArgGlyValValGlyProAlaPhe 103
|||||
685 TCGACCCCGCGCGCCACACGAACTGGCTGCTGACCGGAGATTTC 734
104 SerProSerAlaLeuArgArgLeuGluProValIleAlaGlyThrValAs 120
|||||
735 TCACCTCTGATCGTACCGGCTGCGCCCGCATCGCCGAACTCGGAG 784
120 PASpLeuLeuArgProAlaLeuAlaArgGlyAlaMetAspValValAspG 137
|||||
785 CGAACTCTGGACCGGCTCGGAGCAGACACCGCGCGCGGATCTCGAGG 834
835 GTTCGCGCGCGCGCTCCCGCATCTCGCATCTCGCATCTCGGATC 884
137 IuLeuAlaIleProLeuAlaLeuArgAlaValLeuGlyLeuLeuGlyLeu 153
|||||
154 .ProAlaAlaAsp.....TrpGlyAlaValGlyArgTrpSer..ArgAs 167
|||||
885 CCGCGCGGAGGAGCAGACACCTGTGGCGGCCAACCGGCGCTTCAGGA 934
167 pValGlyArgThrLeuAspArgGlyAlaSerAlaGluAspMetArgArg 184
```

```
935 GGGCGGACACACGCTC.....GGCGCGCGGCCAGCGGTACGACAGG 975
184 IyHisAlaAlaIleAlaGluPheAlaAspTyrValGluArgAlaLeuAla 200
976 CCGAGCGCGGCTCCCGAGAGTTACCGCGTAATTCCCGCGAGAGGTGGAC 1025
201 ArgArgArgArgGluGlyGlyGlyAspLeuLeuAlaLeuMetLeuAspAl 217
1026 CGGCGCGCGGCGGAGCAGACCGCGGAGATCTGCTACCTCTCTCGCGCGC 1075
217 aHisAspArgGlyLeu...MetSerArgAsnGluIleValSerThrValY 233
1076 CCGGAGCAGCGGATACCGGTCACCGGACGAGGAGCATGCTGCGGACCTG 1125
233 aThrPheIlePheThrGlyHisGluThrValAlaSerGlnValGlyAsn 249
1126 TCATCTGCTCACCGCGCGCAGAGACACCACTGCTGCGGAGG 1175
250 AlaValIleSerLeuLeuAlaHisProAspGlnLeuAspLeuLeuArgAr 266
1176 GGGTCTCTACCTGCGCGCGCCACCTGACCTCTCTGACGAGCTGGCGAC 1225
266 gArgProAspLeuLeuAlaGlnAlaValGluGluCysLeuArgTyrAsp 283
1226 CACACCGGAGTGCACACCGCGCGCTCGAAGACTGATGCGTACGACAC 1275
283 roSerValGlnSerAspThrArgGlnLeuAspValAlaGluLeuArg 299
1276 CGCGCGTGCAGCGGTGACGCGCTGGCGCTACGAGACATCGGCTCGGC 1325
300 GlyArgArgLeuArgArgAspAspValValValLeuAlaGlyAlaAl 316
1326 GACACAGACATCCCGCGCGGACCGCGGTGTCGCGCTGCGGCTCGGC 1375
316 aAsnArgAspProArgArgTyrAspArgProAspAspPheAspIleGlu 333
1376 GAACGCGGACCGCGCGCTTCGCGCTCCGACGTGGAGGTCCACAC 1425
333 rGAspProValProSerMetSerPheGlyAlaGlyMetArgTyrCysLeu 349
1426 GCGCGCGCGCAGACGAGAGTGGCTTCGCGCTCGAATCCACTACGCTTC 1475
350 GlySerTyrLeuAlaArgThrGlnLeuArgAlaAlaValAlaAlaLeuAl 366
1476 GGGCGACCGCTGCGCGCGCGCGAGGCGGATCGGTGAGGGCCCTG.. 1523
366 argLeuProGlyLeu...ArgLeuGlyCysAlaSerAspAlaLeuAlAT 382
1524 ...CTGGACGGCATCCCGCGCTCGCGGAGGCGCCACGAGGTGAGT 1569
382 yrcGlnProArgThrMetPheArgGly.....LeuAlaSerLeu 394
1570 AGCGCGACAGATGCTCTTCACAGGCGCGAMCGGCTCTCTCGACCTG 1619
395 ProIleAlaPheThrPro 400
1620 CCGGAMGCCACGPDCCC 1637
```



```
XX Bower SG, Perkins JB, Pero JG, Yocum RR;
XX WPI: 1995-053684/08.
XX Biotin genes, and constructs derived from Bacillus subtilis - for
PT improved production of recombinant biotin or biotin precursor for
PT use in e.g. dietary supplements
XX
XX Example II: Fig 14; 75pp: English.
XX
XX AA081792 is the B. subtilis biotin operon and flanking sequences,
CC as part of an expression vector it can be used in the recombinant
CC production of biotin (or biotin precursor protein). The biotin can
CC be used as a dietary additive in animal feeds, and as a vitamin
CC supplement for human consumption. Biotin is also useful as a
CC reagent for research, and diagnostic procedures.
XX
SQ Sequence 8478 BP; 2432 A; 1748 C; 2088 G; 2205 T; 5 other;

alignment_scores:
Quality: 609.00 Length: 396
Ratio: 2.316 Gaps: 5
Percent Similarity: 66.414 Percent Identity: 33.586

alignment_block:
US-09-724-797-36 x AA081792 ..

Align seq 1/1 to: AA081792 from: 1 to: 8478

8 AlphaheasprothrAspAlaaspValArgArgasProtyrProserTy 24
||| : : : : : : : : : : : : : : : : : : : : : : : :
5493 GCATCGCACTGCATCTTCTGAGTTTGTGAAAACCCATATCTTTTAA 5542
24 HistTrpleuAlarHisasproValHisArgGlyAla.....H 38
||| : : : : : : : : : : : : : : : : : : : : : : : :
5543 CGACACATGCGAGCTGTTCACTATCTATTAAGGAGTTTCTTAAAT 5592
38 IsArgValTrpTyValSerArgPheAlaaspValArgAlaValLeuGly 54
: : : : : : : : : : : : : : : : : : : : : : : :
5593 ACCCGGCTGGATGTCACAGATATGAGAAACGGCTGCTATTTTTAAA 5642
55 AspGluArgPhe...AlaArgThGlyLeaArgPheTrpThrAspLe 70
||| : : : : : : : : : : : : : : : : : : : : : : : :
5643 GATCGGAGATTCAAAGTCGCGACCCCGCTGCTGAGAGCTCAACCAATA 5692
70 uValGlyProGlyLeuAlaGluLeuValGlyAspLeuLeuPheG 87
||| : : : : : : : : : : : : : : : : : : : : : : : :
5693 TCAGGAC.....CTTTCACATGTGCAGAAATCAATGATGCTGTTTC 5733
87 InaAspGluProAspHisGlyArgPheArgGlyValValGlyProAlaPhe 103
: : : : : : : : : : : : : : : : : : : : : : : :
5734 AGAACCCCTGATCATAGACGATTGCGGACGCTGCCAGGAGGCGCTTT 5783
104 SerProSerAlaLeuArgArgLeuGluProValIleAlaGlyThrValAs 120
: : : : : : : : : : : : : : : : : : : : : : : :
5784 ACCCGGACGACAGACAGATTATACGCGGTATATCATTTGAAGACTGCCA 5833
120 PasPheLeuArgProAlaLeuAlaArgGlyAlaMetAspValValaspG 137
||| : : : : : : : : : : : : : : : : : : : : : : : :
5834 TCATTTCCTTATCATCAAGTCAAGTAAAAAAAGATGAGTGCATTTCCG 5883
137 LuLeuAlaTyProLeuAlaLeuArgAlaValLeuGlyLeuLeuGlyLeu 153
: : : : : : : : : : : : : : : : : : : : : : : :
5884 ACTTTCCTTTCTTTAGCAGTTTGTTCATACCTAATATATAGGCTGA 5933
154 ProAlaAlaAspTrpGlyAlaValGlyArgTrpSerArgAspValGlyAr 170
||| : : : : : : : : : : : : : : : : : : : : : : : :
5934 CCGGAGGAAGATAGGAGCAATTAAAGAGATGGGTGCGACTGCATTTCA 5983
170 gThrLeuAspArgGlyAlaSerAlaGluAspMetArgArgGlyHisAla 187
: : : : : : : : : : : : : : : : : : : : : : : :
```

```
5984 AACGATTGATTTTACCCGCTCAAGAAAGCATTTAAGAGGCAATATTA 6033
187 laIleAlaGluPheAlaAspTyValGluArgAlaLeuAlaArgArgArg 203
: : : : : : : : : : : : : : : : : : : : : : : :
6034 TGGCTGTGCAAGGCTATGCGATATTTCAAGAGCTGATTTCAAAAGAA 6083
204 ArgGluGlyGlyLuAspLeuAlaLeuMetLeuAspAlaHisAspAr 220
||| : : : : : : : : : : : : : : : : : : : : : : : :
6084 CCGCACCCCTCAACAGATATGATCAGCATGCTGTGAAGGAGGAGAA 6133
220 gGlyLeuMetSerArgAsnGluIleValSerThyValValThrPheIleP 237
: : : : : : : : : : : : : : : : : : : : : : : :
6134 GGAATAAGCTGACGAGAAAGAGCGGCAATCTAGCTGATATTCGGGCA 6183
237 heThrGlyHisGluThrValAlaSerGlnValGlyAsnAlaValLeuSer 253
: : : : : : : : : : : : : : : : : : : : : : : :
6184 TCGCCGCAATAGACACACGCTCAATCTCATAGCAATTCATCTTTTGT 6233
254 LeuLeuAlaHisProAspGlnLeuAspLeuLeuArgArgArgProAspLe 270
||| : : : : : : : : : : : : : : : : : : : : : : : :
6234 CTGCTGCAGCATCCAGAACAGCTTTGAACAGAGAGAAATCCAGATCT 6283
270 uLeuAlaGlnAlaValGluGluCysLeuArgTyArgProSerValGlns 287
: : : : : : : : : : : : : : : : : : : : : : : :
6284 TATTGTACCGCAGCTCGAGAAATGTTTACGCTATGAAGCCCGCAGCAA 6333
287 eArgAspThrArgGlnLeuAspValAspValGluLeuArgGlyArgArgLeu 303
: : : : : : : : : : : : : : : : : : : : : : : :
6334 TGACAGCCAGAGTGGCGTCAAGAGATATGACATCTCGCGGGTGACAGTC 6383
304 ArgArgAspAspValValValLeuAlaGlyAlaAlaAsnArgAspPr 320
||| : : : : : : : : : : : : : : : : : : : : : : : :
6384 CGTCAGAGAGAACAAAGCTATCTTTGTTAGAGCGGCTAATCGAGACC 6433
320 oArgArgTyTrpAspArgProAspAspPheAspIleGluArgAspProValP 337
: : : : : : : : : : : : : : : : : : : : : : : :
6434 TACATATTTACGAAACCCGATGCTTCGATATATCGAAGATGCTTAATC 6483
337 roSerMetSerPheGlyAlaGlyMetArgTyCysLeuGlySerTyArgLeu 353
|| : : : : : : : : : : : : : : : : : : : : : : : :
6484 CGCATTTTCATTTGCGGCATCGCATCATGTTGCTTAGGCTCTCGCGTG 6533
354 AlaArgThGlnLeuArgAlaAlaVal...AlaAlaLeuAlaArgLeuPr 369
||| : : : : : : : : : : : : : : : : : : : : : : : :
6534 GCACGATTAGAACCGCAAAATGCGATTACACTTCTTCGACCGAATGCC 6583
369 oGlyLeuArgLeuGlyCysAlaSerAspAlaLeuAlaTyArgInProArgT 386
||| : : : : : : : : : : : : : : : : : : : : : : : :
6584 CACCCCTTAATCTT.....GCGGATTTGAATGCGCGTATCGCGGCTTT 6627
386 hrMetPheArgGlyLeuAlaSerLeuProIleAlaPhe 398
||| : : : : : : : : : : : : : : : : : : : : : : : :
6628 TTGGATTTTCGGGCGCTTGAGAGCTCCGCGTACTTTT 6665

seq_name: /SID1/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.AAA75635
seq_documentation_block:
ID AAA75635 standard; DNA; 5970 BP.
XX
XX AAA75635;
XX
XX 22-JAN-2001 (first entry)
XX
XX Nucleotide sequence of ORF12 which encodes a transcriptional activator.
DE Narbonolide synthase; polyketide synthase gene; narbonolide polyketide;
XX antibiotic C12-hydroxylase; pick; desosamine biosynthesis;
KW desosaminyl transferase enzyme; ketoide; beta-glucosidase enzyme;
KW picromycin biosynthesis; ss.
XX
XX Streptomyces venezuelae.
OS
XX US6117659-A.
PN
```

```

XX 12-SEP-2000.
XX
XX
XX 27-MAY-1999; 9905-0320878.
XX
XX 28-MAY-1998; 9805-0087080.
PR 22-SEP-1998; 9805-0100880.
PR 08-FEB-1999; 9905-0119139.
PR 20-MAY-1999; 9905-0134990.
PR 30-APR-1997; 9705-0846247.
PR 06-MAY-1998; 9805-0073538.
PR 28-AUG-1998; 9805-0141908.
XX
PA (KOSA-) KOSAN BIOSCIENCES INC.
XX
PI Ashley G, Betlach MC, Betlach M, Tang L, McDaniel R;
DR WPI: 2000-610844/58.
XX
XX New recombinant pick hydroxylase gene of Streptomyces venezuelae useful
PT for converting ketolides to antibiotics and as antibiotics and
PT intermediates in the synthesis of compounds with pharmaceutical value
XX
PS Disclosure; Columns 41-44; 117pp; English.
XX
CC The present sequence is used to produce the recombinant DNA compounds
CC of the invention. The specification describes a recombinant DNA compound
CC expressing recombinant polyketide synthase genes in host cells for the
CC production of narbonolide, narbonolide derivatives and polyketides that
CC are useful as antibiotics and as intermediates in the synthesis of
CC compounds with pharmaceutical value. The DNA compounds may also encode
CC a C12-hydroxylase (Pick), desosamine biosynthesis and desomycinyl
CC transferase enzymes (useful for conversion of ketolides to antibiotics),
CC and the beta-glucosidase enzyme (involved in picromycin biosynthesis).
CC These compounds are also useful for increasing the antibiotic activity
CC of a compound relative to the unhydroxylated compound. The recombinant
CC host cells are useful as genetic systems that allow rapid engineering
CC of the narbonolide polyketide synthase. These would be valuable for
CC creating novel ketolide analogs for pharmaceutical applications.
XX
SQ Sequence 5970 BP; 806 A; 2219 C; 2142 G; 800 T; 3 other;

Alignment_scores:
Quality: 538.50 Length: 398
Ratio: 2.216 Gaps: 8
Percent Similarity: 61.055 Percent Identity: 34.673

Alignment_block:
US-09-724-797-36 x AAA75635 ..

Align seg 1/1 to: AAA75635 from: 1 to: 5970

15 AspyAAlatgaTgAsPrToTyTProSeTYuNHistTrpleuDeuArgHisas 31
||| |||||||||::: ||| :::
1422 GATTTCGGGCGGCATCCGTATCAGCTACGCCGAGATCGTGCGCGAGG 1471
|||:::|||||
31 ProValHistArg.....GLyAlahisArgValTrpTrytAls 44
|||:::|||||
1472 TC CGGCCCAACCGGCTGGCGCACCCCGGAGGGGAGGAGCAAGTGGTGCTGGT 1521
44 eratgPheAlaAspValArgAlaValLeuGlyAsrGluArgPheAlaArg 60
::: |||||||||::: ||| |||:::
1522 TC GGCTACGACCGGGCGGGCGGTCTCTCCGCATCCCGCGTTACAGCAAG 1571
61 ThrGlyIleArgArgHeTrpThrAspLeuValGlyProGlyLeuLeuAl 77
|||::: ||| ||
1572 GAC.....TGCGCAACTCCACGACTCCCTGACCGAGGC 1606
77 aglu...lIeValGlyAspIleIleLeuPheGlnAspGluProAspHisG 93
||| |::: ||| ||| |||
1607 CGAGCGCGCGCTCAACCAACAATGCTGGAGTGAACCGCGCGGAGCA 1656

```

93	lyrArgLeuArgGlyValValGlyProAlaPheSerProSerAlaLeuArg	109
1657	CCGGCGCTGGCAGACGCTGTGGCCCGCTGATTCACCATGGCCGGGTGCAG	1706
110	ArgLeuGluProValIleAlaGlyThrValAspAspLeuLeuArgProAl	126
1707	TTGTGTGGCGCCCGGGTCCAGAGATCGTCGACGGGCTGTGGAGCCGAT	1756
126	AlaLeuAlaArg.....GlyAlaMetSerPalaValAspGluLeuAlaArgP	141
1257	GCTGGGGGGCCCGACGGCCCGCGCATGTGATGGAGTCCCTGGCTGTGC	1806
141	roLeuAlaLeuArgAlaValAlaLeuGlyLeuLeuGlyLeuProAlaAsp	157
1807	CGCTGGCGGATCACCGTGAATCTCGAATCTCTGGGGTGGCCGAGCCGAG	1856
158	TryGlyAlaValAlaGlyArgTrpSerArgAspValGlyArgThrLeuAspAr	174
1857	CGCGCGCCCTTCCGCGCTGTGGAC.....GAGCG	1885
174	gGlyAlaSerAlaGluAspMetArgArgGlyHisAlaAlaIleAlaGluP	191
1886	CTTGCTGTCCCGGAGATCCCGGCCACGAGCCGACACCCATGGCCGAGA	1935
191	heAlaAspTryValAlaGluArgAlaLeuAlaAlaArgArgArgGluGly	207
1936	TGAGCGGCTATCTCTCCCGGCTATCCAGATCCACACCGGGCAGAGACGCG	1985
208	GluAlaPLeuLeuAlaLeuLeuLeuLeuAspAlaHisAspArg.....GlyLe	222
1986	GAGGACCTGCTCAGCGCGCTCTGGGAGCAGCGACGAGAGAGGCTCCG	2035
222	uMetSerArgAsnGluIleValSerThrValValThrPheIlePheThrG	239
2036	GCTGACCTCCGAGGAGCTCTCGTATGAGCCACATCTGCTGCTGGCGG	2085
239	lyHisGluThrValAlaAspSerGlnValGlyHisAlaValLeuSerLeuLeu	255
2086	GGGACAGACACACGGGTCAATGTGATCGCACAGGAGATTCAGCGCTGTC	2135
256	AlaHisProAspGluLeuAspLeuLeuArgArgArgProAspLeuLeuAl	272
2136	TGCGACCCCGACAGCTGGCCGCTGGGGCGGACATGAGACGCTTGGGA	2185
272	agIlnAlaValGluGluCysLeuArgTryAspProSerValGlnSerAsnT	289
2186	CGCGCGGGTGGAGAGATTTGCTGCTACGAGGCGCGGCGTGAATCCGGA	2235
289	hrArgGluLeuAspValAsp...ValGluLeuAspGlyArgArgLeuArg	304
2236	CTTACCGCTTCCCGGTCCGAGCCCGTGCAGCTGGAGAGGAGCGTCATCCG	2285
305	ArgAspAspValAlaValAlaLeuAlaGlyAlaAlaAsnArgAspProAr	321
2286	GCGGAGACAGACGTCGTGCTGCTCTGGCCGACGCCACGCCGCCGGA	2335
321	gArgTryAspArgProAspAspPheAspIleGluArgAspProValProS	338
2336	GGCGTCTCCCGACCGCAGCCGCTTGACATCCGCGGGAGCACCGCCGCC	2385
338	erMetSerPheGlyAlaGlyMetArgTryCysLeuGlySerTryLeuAla	354
2386	ATCTGCGCTTGGCCGACGAGCATCCACTTCTGTGATCGGCGCCCTTGGCC	2435
355	ArgThrGlnLeuArgAlaAlaVal...AlaAlaLeuAlaArgLeuProGl	370
2436	CGGTTGGAGGCGCGGATCGCGCTCGCGCCCTTCCGACAGCGTCCGGA	2485
370	lyLeuArgLeuGlyCysAlaSerAspAlaLeuAlaValGlnProArgThrM	387
2486	CTTGCCCTTGACAGCTCTCCCGCGGACACTGCTGTATTCGAAACCGGA	2535


```

208  GlnAspLeuLeuAlaLeuMetLeuAspAlaHisAspArg.....GlyLe 222
      |||||||.....:.....:.....:.....:.....:.....:
1986  GAGAGCTGCTCAGCGCTGCTGCGAGCAGCAGCAGCAGCGCTCCG 2035
      :.....:.....:.....:.....:.....:.....:
222  uMeSerArgAnsGluIleValSerThrValIThrPheIlePheThrG 239
      :.....:.....:.....:.....:.....:.....:
2036  GCTACACTCCGAGAGAGCTGCTGCTGATGCGCCACATCTCTGCTGCGG 2085
      :.....:.....:.....:.....:.....:.....:
239  LysIsgLurThrValAlaSerGlnValGlyAsnAlaValLeuSerLeu 255
      |||||||.....:.....:.....:.....:.....:.....:
2086  GGCACGAGACACAGCTCAATCTGATCCCAACGCGCATGTACGCGCTCTC 2135
      :.....:.....:.....:.....:.....:.....:
256  AlaHisProAspGlnLeuAspLeuLeuArgArgArgProAspLeuLeu 272
      :.....:.....:.....:.....:.....:.....:
2136  TCGACACCCGCGACGCTGCGCCGCTGCGGCGCAGCATGACGCTCTTGA 2185
      :.....:.....:.....:.....:.....:.....:
272  aGlaAlaValGlnGluCysLeuArgTyrAspProSerValGlnSerAsn 289
      |||||||.....:.....:.....:.....:.....:.....:
2186  CGGCGCGGTGAGAGAGATGTTGCCCTACGAGGCGCCGCTGGAATCCGCG 2235
      :.....:.....:.....:.....:.....:.....:
289  hrArgGlnLeuAspValAsp...ValGlnLeuArgGlyArgArgLeuArg 304
      ||.....:.....:.....:.....:.....:.....:
2236  CCTACCGCTTCCGCTGCGAGCGCTGACCTGACGCGACGCGCTCATCCG 2285
      :.....:.....:.....:.....:.....:.....:
305  ArgAspAspValValValValLeuAlaGlyAlaAlaAsnArgAspProAr 321
      :.....:.....:.....:.....:.....:.....:
2286  GCCGGGTGACAGGTCCTGCTGCTGCGCGACCGCCACCGCACCCCGGA 2335
      :.....:.....:.....:.....:.....:.....:
321  gArgTyrAspArgProAspAspAspPheAspIleGluArgAspProValPro 338
      ||.....:.....:.....:.....:.....:.....:
2336  GCGCTTCCGCGAGCCGCGACCGCTTGCATCTGCGCGGAGCAGCCGCGGCC 2385
      :.....:.....:.....:.....:.....:.....:
338  erMetSerPheGlyAlaGlyMetArgTyrCysLeuGlySerTyrLeuAla 354
      :.....:.....:.....:.....:.....:.....:
2386  ATCTGCGCTTCCGCGACGCGCATCTCTGCTGATCGCGCGCCCTTGCGCC 2435
      :.....:.....:.....:.....:.....:.....:
355  ArgThrGlnLeuArgAlaAlaVal...AlaAlaLeuAlaArgLeuProG 370
      ||.....:.....:.....:.....:.....:.....:
2436  CGGTTTGAGGCCCGCGATCGCGCTGCGCGCTTCTCGACACGCTCCGCGGA 2485
      :.....:.....:.....:.....:.....:.....:
370  yLeuArgLeuGlyCysAlaSerAspAlaLeuAlaTyrGlnProArgThrM 387
      ||.....:.....:.....:.....:.....:.....:
2486  CTTGCGCTTGGAGCGTCTCCCGCGGGAACGTGTGTGATCGAACCAGCGGA 2535
      :.....:.....:.....:.....:.....:.....:
387  etPheArgGlyLeuAlaSerLeuProIleAlaPheThrProGly 401
      :.....:.....:.....:.....:.....:.....:
2536  TGATCCGCGGCGCTCAAGGCCCTGCGGATCGCTGCGCGCGAGGA 2579
      :.....:.....:.....:.....:.....:.....:

seq_name: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2000.DMT:AAZ87301
seq_documentation_block:
ID   AAZ87301 standard; DNA; 1251 BP.
XX
XX   AAZ87301;
AC
XX
XX   05-JUN-2000 (first entry)
DE
XX
XX   S. venezuelae macrolide biosynthetic gene pikC, SEQ ID NO:38.
KW
XX   Desosamine biosynthesis; macrolide; polyketide; methymycin; pikromycin;
KW   neomethymycin; narbomycin; polyhydroxyalkanote monomer synthase;
KW   biopolymer; antibiotic; chemotherapeutic; immunosuppressant; asthma,
KW   chronic obstructive pulmonary disease; respiratory inflammation;
KW   hypercholesterolemia; crop protection agent; ds.
OS
XX   Streptomyces venezuelae ATCC15439.
FH
XX   Key location/Qualifiers
FT
XX   CDS 1..1251
FT
XX   /*Lag= a
FT   /product= "pikC"
FT   /transl_except= (pos:307..309, aa:Leu)
FT

```

```

XX
XX   WO20000620-A2.
XX
XX   06-JAN-2000.
XX
XX   25-JUN-1999; 99WO-US14398.
XX
XX   26-JUN-1998; 98US-0105537.
XX
XX   (MINU ) UNIV MINNESOTA.
XX
XX   Sherman DH, Liu H, Xue Y, Zhao L;
XX
XX   WPI: 2000-160679/14.
XX
XX   P-PSDB: AAY77196.
XX
XX   Desosamine and macrolide biosynthetic gene clusters, useful for, e.g.
XX
XX   synthesis of methymycin and pikromycin -
XX
XX   Claim 15; Page 428-429; 438pp; English.
XX

```

The invention relates to an isolated and purified nucleic acid segment comprising a desosamine biosynthetic gene cluster, a fragment or its biologically active variant, where the nucleic acid sequence is not derived from the eryC gene cluster of *Saccharopolyspora erythraea* or *Streptomyces antibioticus*. The invention also relates to a macrolide biosynthetic gene cluster, or fragments thereof. The macrolide biosynthetic gene cluster encodes proteins which synthesise methymycin, pikromycin, neomethymycin, narbomycin or a combination of these compounds. Recombinant or augmented cells comprising the desosamine and/or macrolide biosynthetic gene clusters are useful for the production of biologically active macrolides. The macrolide biosynthetic proteins are useful for synthesis of methymycin, pikromycin, neomethymycin and narbomycin. The alternative termination of polyketide synthesis may be useful to prepare novel antibiotics and polycyetroxalcanate (PHA) monomers. The compounds produced by the recombinant host cells are useful as biopolymers, e.g., in packaging or biomedical applications, to engineer PHA monomer synthases or to prepare biologically active agents, such as chemotherapeutics, immunosuppressants, agents to treat asthma, chronic obstructive pulmonary disease as well as other diseases involving respiratory inflammation, cholesterol-lowering agents or macrolide-based antibiotics which are active against a variety of organisms, e.g., bacteria, including multi-drug resistant pneumococci and other respiratory pathogens, as well as viral parasitic pathogens, or as crop protection agents (e.g., fungicides or insecticides) via expression of polyketides in plants. Sequences AAZ87295-287302 represent macrolide biosynthetic genes from *Streptomyces venezuelae* ATCC 15439, which encode proteins AAY77190-y77197.

Sequence 1251 BP; 171 A; 484 C; 413 G; 183 T; 0 other;

alignment_scores: Length: 398
Quality: 531.50
Ratio: 2.196
Percent Similarity: 60.804 Percent Identity: 34.422

alignment_block:
US-09-724-797-36 x AAZ87301 ..

Align seg 1/1 to: AAZ87301 from: 1 to: 1251

```

15  AspValArgArgAspProTyrProSerTyrHisTrpLeuArgHisAs 31
      |||.....:.....:.....:.....:.....:.....:
67  GATTTCGCGCGCGATCCGTAACGAGCTACGCGAGACTGCGTCCAGAGG 116
      :.....:.....:.....:.....:.....:.....:
31  pProValHisArg.....:.....:.....:.....:.....:.....:
      :.....:.....:.....:.....:.....:.....:
117  TCGCGCCACCGGCTGCGGACCGCCGAGGGGAGCAGAGTGTGCTGTGCG 166
      :.....:.....:.....:.....:.....:.....:
44  erArgPheAlaAspValArgAlaValLeuGlyAspGluArgPheAlaArg 60
      :.....:.....:.....:.....:.....:.....:

```

```

167 TCGGCTACGACCGGCGCGGCGGTCTCGCCGATCCCGCGGTGACGAG 216
61 ThrGlyIleArgArgPheTrpThrAspLeuValGlyProGlyLeuAla 77
217 GAC.....TGGCGCAACTCCACGACTCCCTGACCGAGGC 251
77 aglu...lleValGlyAspIleIleLeuPheGlnAspGluProAspHisG 93
252 CGAGGCCCGCGCTCAACACACACATGCTGGAGTCCGACCGCGCGGACACA 301
93 lYArgLeuArgGlyValValGlyProAlaPheSerProSerAlaLeuArg 109
302 CCGCGCCCGCGCAAGCTGGTGGCGGTGATCCACATCGCGCGGTGAG 351
110 ArgLeuGluProValIleAlaGlyThrValAspAspLeuLeuArgProAl 126
352 TTGGTGGGCGGCGCGGCTCCAGAGATGCTGCACGGGCTCGTGAGCGCAT 401
126 aLeuAlaArg.....GlyAlaMetAspValValAspGluLeuAlaTyrP 141
402 GCTGGCGGCGCGCGCGACGCGCGCGCATGTGATGAGTCCCTGGCTGGC 451
141 rOlauAlaLeuArgAlaValLeuGlyLeuLeuGlyLeuProAlaAlaAsp 157
452 CGCTGCCGATCACCGTGATCTCCGAACTCTCGGCGGTGCCGACCGGAC 501
158 TrpGlyAlaValGlyArgTrpSerArgAspValGlyArgThrLeuAspAr 174
502 CGCGCGCGCTTCGCGGTCTGGACC.....GACGC 530
174 gGlyAlaSerAlaGluAspMetArgArgGlyHisAlaAlaIleAlaGluP 191
531 CTGTGCTTCCGCGGACGATCCCGCGCGCGCACGACCGCATGGCCGAGA 580
191 heaLaAspTyrValGluArgAlaLeuAlaArgArgArgGluGly 207
581 TGACGGGCTATCTCTCCGCTCATCGATCCCAAGCGCGGCGAGAGCGG 630
208 GluAspLeuLeuAlaLeuMetLeuAspAlaHisAspArg.....GlyLe 222
631 GAGGAACTGCTCAGCGCGCTCGTCCGCGACGAGGACGAGGACGGCTCCG 680
222 uMeSerArgAsnGluIleValSerThrValValThrPheIlePheThrG 239
681 GCTGACCTCCGAGAGCTGCTCGTATGAGCCACATCTCTGCTGCGCG 730
239 lYHisGluThrValAlaSerGluValGlyValAlaValLeuSerLeuLeu 255
731 GCGACGAGACACGCGTCAATCTGATCGCCACACGCGATGATACGGCTGC 780
256 AlaHisProAspGluLeuAspLeuLeuArgArgArgProAspLeuAla 272
781 TCGACACCCGACGAGTGGCGCGCTGGCGCGGACATGACGCTTTGG 830
272 agInAlaValGluGluCysLeuArgTyrAspProSerValGlnSerAsnT 289
831 CGGCGGCGTGGAGAGATGTTGGCTACGAGGCGCGGTGGATCCGCCA 880
289 hrArgGlnLeuAspValAsp...ValGluLeuArgGlyArgGluArg 304
881 CTRACCGCTTCGCGGTGAGACCGCTGACTGAGACGCGACGGCTCATCCG 930
305 ArgAspAspValValValValLeuAlaGlyAlaAlaAsnArgAspProAr 321
931 GCGGCTGACACGGTCTCGTCTGCGGCGAGCGCCACCGCACCCCGCA 980
321 gaArgTyrAspArgProAspAspPheAspIleGluArgAspProValProS 338
981 GCGCTTCCCGGACCCGACCGCTTCGACATCCGCGGACACCGCGCGCC 1030
338 erMeSerPheGlyAlaGlyMetArgTyrCysLeuGlySerIleAla 354
1031 ATCTCCGCTTCGCGCACGCGCATCTTCTGCACTCGCGCGCGCTTGGCC 1080

```

```

355 ArgThrGlnLeuArgAlaAlaVal...AlaAlaLeuAlaArgLeuProG 370
1081 CGGTTGGAGGCGCGGATCGCGCTCCGCTTCTCGAACGCTGCGCCGA 1130
370 yLeuArgLeuGlyCysAlaSerAspAlaLeuAlaTyrGlnProArgThiM 387
1131 CCTCGGCCCTTGAGCTTCCCGCGCGCAACTCGTGTGTCCGAAACCCGA 1180
387 eUrPheArgGlyLeuAlaSerLeuProIleAlaPheThrProGly 401
1181 TGATTCGCGGCTCAAGCGCCCTCGCATCCGCTGGCGGCGGAGGA 1224

seq_name: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.AAC5790
seq_documentation_block:
ID AAC5790 standard; DNA; 1233 BP.
XX
AC AAC5790;
XX
DT 19-JAN-2001 (first entry)
XX
DE Hydroxylase orf3 involved in MC biosynthesis.
XX
KW Mitomycin; biosynthesis; mitosane ring system; antibiotic; anti-cancer;
KW anti-inflammatory; immune-enhancer; immunosuppressant; asthma;
KW chronic obstructive pulmonary disease; respiratory inflammation;
KW fungicide; pesticide; ds.
XX
OS Streptomyces lavendulae.
XX
PN WO200053737-A2.
XX
PD 14-SEP-2000.
XX
PF 10-MAR-2000; 2000WO-US06394.
XX
PR 12-MAR-1999; 99US-026965.
XX
PA (MNU) UNIV MINNESOTA.
PA (SHER/) SHERMAN D H.
PA (MAOY/) MAO Y.
PA (VARO/) VAROGLU M.
PA (HEMM/) HE M.
PA (SHEL/) SHELDON P C.
XX
PI Sherman DH, Mao Y, Varoglu M, He M, Sheldon PC;
XX
DR WPI; 2000-601980/57.
XX
PT Novel nucleic acid molecule comprising mitomycin biosynthetic gene
PT cluster useful for cloning mitomycin biosynthetic genes for elucidating
PT the molecular basis of mitosane ring system biosynthesis
XX
PS Example 1; Page 255; 399pp; English.
XX
CC This invention relates to isolated and purified nucleic acid molecules
CC from the mitomycin biosynthetic gene cluster. Mitomycins are a group of
CC natural products that contain a variety of functional groups, including
CC amino benzquinone and axiridine ring systems. The S. lavendulae
CC mitomycin biosynthetic gene cluster comprises 47 mitomycin genes
CC spanning 55kb of DNA. The invention includes an expression cassette
CC comprising a mitomycin biosynthetic gene operably linked to a promoter,
CC and host cells transformed with the cassette. The nucleotide, and protein
CC sequences and the transformed host cells of the invention result in
CC antitastmatic, antiinflammatory, cytostatic, immunomodulatory, and
CC antibiotic activities. The nucleotide sequences are used to elucidate the
CC molecular basis for the biosynthesis of the mitosane ring system, as well
CC as to engineer the biosynthesis of novel natural products, e.g.
CC antibiotics, anti-inflammatory agents, anti-cancer agents,
CC immune-enhancers, immunosuppressants, agents to treat asthma, chronic
CC obstructive pulmonary disease as well as other disease involving
CC respiratory inflammation, or cholesterol-lowering agents or as crop

```

CC protection agents (e.g. fungicides or insecticides) as well as
 CC biopolymers, e.g., in packaging or biomedical applications, or to engineer
 CC PHA monomer synthases. Sequences AAC55782-C55881, AAC55815-C55849 and
 CC AAB2485-B23242 represent mitomycin biosynthetic gene cluster DNA
 CC sequences and encoded proteins. Sequences AAC55812-C55814,
 CC AAC55850-C55856 and AAC55862-C55869 represent PCR primers used in the
 CC cloning of the mitomycin biosynthetic genes.
 XX

SQ Sequence 1233 BP; 165 A; 479 C; 421 G; 168 T; 0 other;

alignment_scores:

Quality: 508.50 Length: 402
 Ratio: 2.067 Gaps: 8
 Percent Similarity: 61.194 Percent Identity: 33.085

alignment_block:
 US-09-724-797-36 x AAC55790 ..

Align seg 1/1 to: AAC55790 from: 1 to: 1233

```

11 ProThrAspAlaSerPValArgArgAspProTyrProSerTyrHisTrpLe 27
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
64 CCCTTCACAGCCCCCGACCGGGGTGAGACCCGACCGGTACTGGAGCCGCT 113
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
27 uLeuArgHisAspProValHisArg.....GlyAlaHisArgV 40
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
114 GCGCCGGAGCGCGCTGCACGCTGCACGCTGCCTGACGGCGGCGAGG 163
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
40 alrPrrValSerArgPheAlaSerPValArgAlaValLeuGlyAspGlu 56
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
164 CGTGGCTCGCGACCGCTATCAGAGGTGCGCGGCTCTCGCGACCGC 213
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
57 ArgPheAlaArg.....ThrglyIleArgArgPheTrpTh 68
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
214 AGGTTCTCCCGGACGTGCGCGTCCGCGCGCTCGCGCTCTCTCC 263
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
68 rAspLeuValGlyProGlyLeuLeuAlaGluIleValGlyAspIleIleL 85
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
264 GCACAGCGCCCGCG.....GACGCGCGTCC 289
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
85 euPheGlnAspGluProAspHisGlyArgLeuArgGlyValValGlyPro 101
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
290 TGAAGCTGAGGGGCCGACGACGCGGCTGCGCGTGTGCGGAG 339
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
102 AlaPheSerProSerAlaLeuArgArgLeuGluProValIleAlaGlyTh 118
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
340 GTCTTCACGCGCGCGCGGTGAGACATGCGTCCGCTCATCCAGCGAC 389
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
118 rValAspAspLeuLeuArgProAlaLeuAlaArgGly..AlaMetAspV 134
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
390 CCGCGACGAGCTCTCGACCGCATGAGAGATGGGCGCGCGCGGAGAC 439
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
134 alValAspGluLeuAlaTyrProLeuAlaLeuArgAlaValLeuGlyLeu 150
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
440 TGGTCGAGGACTCTCCCTGCGCGTGCATGATTCGCGAGCGT 489
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
151 LeuGlyLeuProAlaAlaAspTrpGlyAlaValGlyArgTrpSerArgAs 167
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
490 CTGGCGCTGCGCGCGCGAGACCGCATGCTTGCCTGTGTCGACGC 539
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
167 pValGlyArgTrpLeuAspArgGlyAlaSerAlaGluAspMetArgArg 184
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
540 GCTGCTGACGACC.....ACCGCGACACCCCGCGCGAG 574
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
184 lyHisAlaAlaIleAlaGluPheAlaAspTyrValGluArgAlaLeuAla 200
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
575 TGGCGGACTCATGATGACATGACAGACTACTCGCGCGGCTCGCGG 624
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
201 ArgArgArgArgGluGlyGlyGluAspLeuAlaLeuMetLeuAspAl 217
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
625 CAGCGCGCGGTGCGCGCGACCGCGGACCTGATCGCTCCCTGTCGACCC 674

```

```

217 aHisAsp...ArgGlyLeuMetSerArgAsnGluIleValSerThrValV 233
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
675 GCGCGACGAGAGACACACTACCGAGGCGGAGCTGTGCGCGTGGCGG 724
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
233 alrPheIlePheThrGlyHisGluThrValAlaSerGluValGlyAsn 249
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
725 AGGCATCTCTCATCGCGGCTACGAGACCTCGCGAGCATCCCAAC 774
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
250 AlaValLeuSerLeuLeuAlaHisProAspGluLeuAspLeuArgArg 266
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
775 TCCCTCTACGCTCTTCCGCCACCGGACGCTGGAGCGGATCGAG 824
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
266 gatArgProAspLeuLeuAlaGluAlaValGluGlyCysLeuArgTrpAsp 283
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
825 CGACACGACATCATCCCGGACCGCGTGAAGACATGCTGCTCCGTCGC 874
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
283 ro.....SerValGlnSerAsnThrArgGluLeuAspValAspValGlu 297
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
875 CCATCGGACCGGTGAGCGCTTCCCGTACGCGCCACCGAGACGTCGAG 924
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
298 LeuArgGlyArgArgLeuArgArgAspAspValValValLeuAlaGlu 314
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
925 CTCGGGGGAGTCTCTGTACAGGCGCGGAGACGCTGTGCTCGATGAG 974
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
314 yAlaAlaAsnArgAspProArgArgTrpAspArgProAspAspPheAspI 331
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
975 GCGCGCACACCGGACCGCGAGCTGTTCAGGACCCCGAGCTGAGACC 1024
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
331 leGluArgAspProValProSerMetSerPheGlyAlaGlyMetArgTrp 347
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1025 TCGCGGGGCGCGGATTCGACACTGTGGCTTCGCGGCGGAGCACACC 1074
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
348 CysLeuGlySerTyrLeuAlaArgTrpGlnLeuArgAlaAlaValAlaAl 364
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1075 TGCTTGAGCGCCCACTGAGCGCGGAGCTCCAGATCCATCCATCCAGC 1124
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
364 aLeu...AlaArgLeuProGlyLeuArgLeuGlyCysAlaSerAspAlaL 380
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1125 GCTGTTCGCGAGATACCCCGCGCTGCGGCTGCGCGGAGAGAGACC 1174
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
380 euAlaTyrGlnProArgTrpMetPheArgGlyLeuAlaSerLeuProle 396
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1175 TCTGCTGAGAGAGAGGCGCTGATGTCTCGCGCATCGACACCATGCGG 1224
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
397 AlaPhe 398
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1225 ACCGTGG 1230

```

seq_name: /STDS1/gcgdata/geneseq/geneseq-emb1/NA1997.DAT:AA70153

seq_documentation_block:

ID AA70153 standard; DNA; 6085 BP.

XX AA70153;

DT 23-JUL-1997 (first entry)

XX

DE S. longisporoflavus staurosporin synthesis gene cluster 6.5kb fragment.

XX

KW staurosporin; gene cluster; indole-carbazole alkaloid; ICA; antibiotic;

KW antiproliferative; platelet aggregation; fungus; yeast;

KW Ca2+/phospholipid-dependent serine/threonine protein kinase; ds.

OS Streptomyces longisporoflavus (strain R19).

XX

XX Key Location/Qualifiers

XX FH 378..1655

XX CDS

XX FT /tag= a

XX FT /label= Gene_1

XX FT /note= "Encodes a protein containing 425 amino acids"

XX FT 1747..2553

XX FT /*tag= b

XX FT /label= Gene_2


```

FT      /note= "Encodes a protein containing 268 amino acids
FT      significantly similar to known S-adenosyl
FT      methionine-dependent methyl transferase"
FT      CDS
FT      2593..4011
FT      /tag= c
FT      /label= Gene_3
FT      /note= "Encodes a protein containing 472 amino acids"
FT      CDS
FT      4013..4999
FT      /tag= d
FT      /label= Gene_4
FT      /note= "Encodes a protein containing 328 amino acids"
FT      CDS
FT      5071..6085
FT      /tag= e
FT      /label= Gene_5_(part)
FT      /note= "Partial sequence for gene 5; full length
FT      protein contains 366 amino acids and is
FT      significantly similar to amino transferase
FT      enzymes, such as the Dnr J protein"
FT
PN      WO9708323-A1.
XX      06-MAR-1997.
XX
XX      19-AUG-1996; 96WO-EP036643.
XX
XX      30-AUG-1995; 95EP-0810534.
XX
XX      (CIBA ) CIBA GEIGY AG.
XX
PI      Bietenhader J, Engel N, Pospiech A, Schupp T, Toupet C;
DR      WPI, 1997-179280/16.
XX
XX      Indole-carbazole alkaloid biosynthesis gene cluster - especially
XX      PT      coding for the antibiotic staurosporin from Streptomyces
XX      longisporoflavus
XX
XX      Claim 7; Page 37-41; 55pp; English.
XX
XX      The present sequence represents the 6.5kb PvulI fragment of Streptomyces
XX      longisporoflavus R19 which is involved in the biosynthesis of
XX      indole-carbazole alkaloids (ICA). The sequence contains five functional
XX      fragments as indicated in the features table. The DNA or a hybrid
XX      vector containing it can be used to prepare an ICA or derivatives and
XX      precursors, either by allowing production in previously incapable
XX      organisms or by improving yields. In particular, the antibiotic
XX      staurosporin can be produced. Staurosporin is known to have inhibitory
XX      activity against fungi, yeasts, and Ca2+/phospholipid-dependent
XX      serine/threonine protein kinases (PKCs). Staurosporin also has
XX      antiproliferative activity and can inhibit platelet aggregation.
XX      The present sequence can also inactivate ICA biosynthesis genes and
XX      can be used in PCR amplification. An advantage of this is that
XX      CC      productively of staurosporin-synthesizing Streptomyces is improved
XX      over natural strains yielding only low concentrations.
XX
XX      Sequence 6085 BP; 882 A; 2374 C; 1992 G; 833 T; 4 other;
XX
XX
XX      alignment_scores:
XX      Quality: 508.50      Length: 407
XX      Ratio: 2.101      Gaps: 11
XX      Percent Similarity: 59.459      Percent Identity: 36.364
XX
XX      alignment_block:
XX      US-09-724-797-36 x AAT70153 ..
XX
XX      Align seg 1/1 to: AAT70153 from: 1 to: 6085
XX
XX      20 ProTyrProSerTyrHisTrpLeuAlaArgHisAspProValHisArg.. 35
XX      |||||  |||||  ::  |||||
XX      441 CCGTACCGGCTACCGGCGGAGCCGCCCGGCTCATTCGAC 490
XX      36 .....GlyAlaHisArgValTrrTyrValSerArgPheAla 48

```

```

491 GCGCGCGCGCCCGGAAAGCTGACACCTACTAGTGTTCACCTACGACG 540
FT      |||  :::::|||||  :::  |
48 spValArgAlaValLeuGlyAspGluArgPheAlaArgTrgIleArg 64
FT      |||||  |||||  :::::|||||  :::::|||||
541 ACGTGTCCGCTCTGCTCCACCGCGGCTTCGCGCCGACAGCC..... 584
65 ArgPheTrpThrAspLeuValGlyProGly..... 74
FT      |||  :::::|||||
585 GCGGTGGCTCCGGCGACACCGCGCCGACACCGCGCCGCTCCGATCC 634
75 .....LeuValGluIleValGlyAspIleIleLeuPheG 87
FT      |||  :::::||||
635 GCGCGAGACCGCGCCCTCGCGACCGTGTGAGAACACTGCTGTCCTCC 684
87 LnsArgLupProAspHisGlyArgLeuArgGlyValValGlyProAlaPhe 103
FT      |||  |||  |||  ::|||  |||||  :::::|||||
685 TCGACCCCCCGCGCCACACCGACTGCGCTCCCTGACCGCGGAGTTC 734
104 SerProSerAlaLeuArgArgGluLupProValIleAlaGlyThrValAs 120
FT      |||||  |||  ::|  |||||  |||||
735 TCACCTCGATCGTCACCGGCTGCGCCCGCATCCGCAACTCGCGGAG 784
120 PASPLeuLeuArgProAlaLeuAlaArgGlyAlaMetAspValAlaSpG 137
FT      :::::|||||  |||||  |||  :::::|||||
785 CGAACTCTGGACCGCGCTCGAGCACACCGCGCCGCGATCTCTCGAGG 834
137 LuLeuAlaTyrProLeuAlaLeuArgAlaValLeuGlyLeuGlyLeu 153
FT      |||  ::|  :::::|||||  :::::|||||
835 GTTCGCGCGCCCTCCCGGATCTCTGCTCCGACCTGCTGGCATC 884
154 ProAlaAlaAsp.....TrpGlyAlaValGlyArgTrpSer..ArgAs 167
FT      |||  |||  |||  |||  |||||  :::::
885 CCCCCCGGAGAGACCAACCTGTGGCGCCCAACGGGTGGCCCTTCAGGA 934
167 PValGlyArgThrLeuAspArgGlyAlaSerAlaGluAspMetArgArg 184
FT      :::::||||  |||||  |||  |||  |||
935 GCGCGCGCACACGCTC.....GCGCGCGCGCCACGATACGACGCGG 975
184 LHisAlaAlaIleAlaGluPheAlaAspTyrValAlaArgAlaLeuAla 200
FT      :::::|||||  |||||  |||  :::::
976 CCGAGCGCGCTCCACGAGATTCACCCCTACTTCGCGGAGAGGTGAC 1025
1026 ArgArgArgArgGluGlyGlyLnsAspLeuAlaLeuMetLeuAspAl 217
FT      |||||  ::|  ::|||  |||||  |||||  |||
201 ArgArgArgArgGluGlyGlyLnsAspLeuAlaLeuMetLeuAspAl 217
1026 CCGCGCGCGCGCGACGACCGCGCATGCTCTCCCTCTCGCGCGC 1075
217 HisAspArgGlyLeu...MetSerArgAsnGluIleValSerThrVal 233
FT      |||||  |||||  |||||  |||||  |||||  |||
1076 CCGGACACCGGATACCGCTAGCGTGGACCGCATCTCGCACCTGCG 1125
233 AlThrPheIlePheThrGlyHisGluThrValAlaSerGlnValGlyAsn 249
FT      ||  :::::  ::|||  |||||  |||||  :::::
1126 TCACATGTGTACCGCGCGCGCACGACACCACTGCTCTCCACGAG 1175
250 AlaValLeuSerLeuLeuAlaHisProAspGlnLeuAspLeuLeuArg 266
FT      |||||  |||||  |||||  |||||  |||||
1176 GCGGTCTCTCCCTGCGCCCACTGACCTCTGAGAGACTCGAC 1225
266 GArgProAspLeuLeuAlaGlnAlaValGluGluGlySerArgTyrAsp 283
FT      |||||  :::::  |||||  |||||  |||||
1226 CACACCGGAGTGCACACCGCGCGCTCGAAGAGTGAAGCGGTACGAC 1275
283 roSerValGlnSerAsnThrArgGlnLeuAspValAlaGluLeuArg 299
FT      ||  |||||  ::|  |||||  |||||  |||
1276 CGCCGCTGACGCGGTGACGCGCTGCGCTGACGAGACATCGGCTCGGC 1325
300 GlyArgArgLeuArgArgAspValValAlaValLeuAlaGlyAlaAl 316
FT      :::::  :::::  |||||  |||||  |||||  |||
1326 GACCAAGACATCCCGCGCGCAACCGGTGTGCGCTCTCGGCTCGGC 1375
316 aAsnArgAspProArgArgTyrAspArgProAspAspPheAspIleGluA 333
FT      |||||  |||||  |||||  |||||  |||||

```


Sequence 50937 BP; 6672 A; 16253 C; 19272 G; 8740 T; 0 other;

FT CDS 41878..43158
FT /tag= h
FT /label= oleG1
FT CDS 43163..44443
FT /tag= i
FT /label= oleG2
FT CDS 44433..45173
FT /tag= j
FT /label= oleM1
FT 45251..46411
FT /tag= k
FT /label= oley
FT 46491..47714
FT /tag= l
FT /label= olep
FT complement (47808..49517)
FT /tag= m
FT /label= oleB

MO200026349-A2.
11-MAY-2000.

XX PD 22-OCT-1999; 99WO-US24478.
XX PF 29-OCT-1998; 98US-0106100.
PR 16-FEB-1999; 99US-0120254.
PA (KOSA-) KOSAN BIOSCIENCES INC.

PL Belach MC, Shah SK, McDaniel R, Tang L;
DR WPI: 2000-365602/31.
P-PSDB; AAY92707, AAY92708, AA92709.

PT Recombinant DNA compound encoding oleandrolide polyketide synthase for
PT synthesizing polyketides comprising a coding sequence for a domain of a
XX loading module or any one of extender modules

XX PS Disclosure; Page 14-26; 86pp; English.

CC This is part of the Streptococcus antibioticus oleandomycin gene cluster
CC The oleandrolide polyketide synthase (PKS), also known as
CC 8,8a-deoxyoleandrolide synthase, is encoded by three open reading frames
CC (ORF), designated oleA1, oleAII and oleAIII. The PKS is a type I
CC "modular" enzyme, where each ORF encodes 2 extender modules and
CC the first ORF also encodes the loading module. Each module is composed
CC of at least a ketosynthase (KS), acyl-transferase (AT) and an
CC acyl carrier protein (ACP) domain. The oleandrolide PKS loading module
CC contains an inactivated KS, called KS-Q, where Q is the abbreviation for
CC glutamine, present instead of the active site cysteine required for
CC activity. The large multifunctional PKS enzymes catalyze the biosynthesis
CC of polyketide macroketones through multistep pathways involving
CC decarboxylative condensations between acylthioesters followed by cycles
CC of varying beta-carbon processing activities. The macroide product of
CC the PKS, 8,8a-deoxyoleandrolide, is further modified by epoxidation and
CC glycosylation to yield oleandomycin, an antibacterial polyketide. The
CC invention concerns an isolated recombinant DNA compound, comprising a
CC coding sequence for a domain of loading module or any one of extender
CC modules 1-4 or 1-6, including an oleandrolide PKS operably linked to a
CC promoter. Also discussed are recombinant oleandrolide PKS in which the
CC module 1 KS domain is inactivated by deletion or other mutation. In
CC particular, the inactivation is mediated by a change in the KS domain
CC that renders it incapable of binding substrate (the KS1-o mutation),
CC rendered by mutation in the codon for the active site cysteine. The
CC oleandrolide PKS is useful for synthesizing polyketides, which are useful
CC as antibiotics and molluscids. Heterologous expression of oleandrolide PKS
CC in host cells such as Streptomyces coelicolor and S. lividans is also
CC made possible. Unmodified oleandrolide compounds can be provided to
CC cultures of Saccharopolyspora erythraea and converted to the
CC corresponding derivatives of erythromycins A-D.

alignment_scores:	Quality:	472.50	Length:	423
Ratio:	1.898	Gaps:	12	
Percent Similarity:	58.865	Percent Identity:	34.043	
alignment_block:				
US-09-724-797-36 x AAA09469 ..				
Align seg 1/1	to: AAA09469	from: 1	to: 50937	
4	AspAlaValThrAlaPheAspProThrAspAlaAspValAlaArgAsp	20		
46524	GAGGGGTACCGGCTAC...CCGTCACACCTGCCGACCGCCTGACCT	46570		
20	oTyrrProSerTyrrHisTrpLeuAlaArgHisAspProValHisArg...	35		
46571	CGACCGGCACTACGGCAACTCGCGCGACGACACCGCTCCACAGGTGC	46620		
36GlyAlaHisArgValTrpTyrValSerArgPheAlaAsp	48		
46621	GCCTGCCCTACGAGGCGACGCGCTGGCTGGTGTACCGCATGTCGAC	46670		
46621	GCCTGCCCTACGAGGCGACGCGCTGGCTGGTGTACCGCATGTCGAC	46670		
46671	GCCCGATGTTCTGGGCGACCTCCGCTTCACACCGCGCGCC...	46715		
65	gPheTrpHisAspLeuValGlyProGlyLeuLeu...AlaGluIleValG	81		
46716ACCGATCCCGACCCCGCGATGTCGCCACCGCGCGGAC	46758		
81	LyAspIleIleuPheGlnAspGluProAsnHisGlyArgGluArgGly	97		
46755	CGGAGGCGCTCTCCCGACGACCGCGGACACCGCGGTGGGCGG	46808		
98	ValValGlyProAlaPheSerProSerAlaLeuArgArgLeuGluProVa	114		
46809	CTGGTGGGCAAGGCTTCACGCGACGCGGGGAGAGATGGCGCCCG	46858		
114	IleAlaGlyThrValAspAspLeuAlaArgProAlaLeuAlaArgGlyA	131		
46859	TGTCGGTCTCCCTGCGACCTCCCTGCGACACATGGTGGCGACGTT	46908		
131	La...MetAspValValAspGluLeuAlaTyrProLeuAlaLeuArgAla	146		
46909	CACCGCGCACTGGTGCAGTTCCTGCGCGCTCCCTCCGTCGGGTC	46958		
147	ValLeuGlyLeuLeuGlyLeuProAlaAlaAspTrpGlyAlaValGlyAr	163		
46959	ATCTGCAACTGCTGGGCTGCCCTTGAGGAC.....	46991		
163	gTrpSerArgAspValGlyArgThrLeuAsp.....	173		
46992CGCACCTGTCGCGACCTTCTGCGAGGCCATGCTCTCGA	47034		
174	..ArgGlyAlaSerAlaGluAspMetArgArgGlyHisAlaAlaIleAla	189		
47035	CCCGGCTACCGCGCGAGATACGCG.....GTCCAGAG	47072		
190	GluPheAlaAspTyrValGluArgAlaLeuAlaArgArgArgGluGlu	206		
47073	GACTTCATGTCATCATGACGAGGCTGGTGGCCGAGGCGCGCGCC	47122		
206	YglYgluAspLeuLeu...AlaIleuMetLeuAspAlaHisAspArgGlyL	222		
47123	CACGAGGACTGCTGGCGGCTCGGCTCGCCACCGACGACGACGAC	47172		
222	euketSerArgAsnGluIleValSerThrValValThrPheIlePheThr	238		
47173	ACCGACCAAGGCGAGATCTCTCAACATGGGGGTGACACCTGTCATCGG	47222		
239	GlyHisGluThrValAlaSerGlnValGluValAsnAlaValIleuSerLeu	255		

```

|||||
47223 GGGCAGAGAGCTGGCTCAACGAGATCCACACCTGCTCCTCTGCT 47272
255 uAlaHisPrcAspGlnLeuAspLeuArgArgArgProAspLeuLeu 272
|||||
47273 GACCGAGCGGAGGCTACGAGTGGCTGGCGGACCGCGGCTGCTGG 47322
272 laGlnAlaValGluGluCysLeuArgTyrAspProSerValGlnSeran 288
|||||
47323 CCGGGGGGTGGAGAGATGCTGGCTGACACACCGCTGGTCCGGCCGGC 47372
289 Thr.....ArgGlnLeuAspValAspValGlnLeuArgGlyArgArg 303
|||||
47373 AGCTTCCTCCGCGCGGCGACCGGAGAGCTGAGCTGAGCACCGTGACCG 47422
303 uArgArgAspAspValValValLeuAlaGlyAlaAlaAsnArgAsp 320
|||||
47423 GCGGGCGGGAGCCCTGCTGCTCCACTTGGCGTGGCCACCGGGAGCG 47472
320 roArgArgTyrArgAspArgProAspAspPheAspPheGlnuArgAspProVal 336
|||||
47473 AGGAGGTCTTGACACCGGACGAGCTGAGACTTCCACCGTGGAGCGCAC 47522
337 ProSerMetSerPheGlyAlaGlyMetArgTyrCysLeuGlySerTyrLe 353
|||||
47523 CCGCACATAGCGTTCGGGCGACGAGCGCCACCATGCTGCGCGCCCACT 47572
353 uAlaArgThrGlnLeuArgAlaAlaValAlaAlaLeuAla...ArgLeuP 369
|||||
47573 GGGCGGACTGGAACTCCAGAGGCGCTGCGCGCTGCTGCGCGGCTTCC 47622
369 roGlyLeuArgGluGlyCysAlaSerAspAlaLeuAlaTyrGlnProArg 385
|||||
47623 CCACCTCGATCTGCGCGAGCGGCTCGCGGACTGGAAGTGAAGCAGAGGC 47672
386 ThrMetPheArgGlyLeuAlaSer.....LeuProI 396
|||||
47673 ATGCTGATCCGCGGACTGGAAGCGCAGATGCTCTGCTGAGCGCGCGC 47722
396 eAlaPheThrProGlyGly 402
|||||
47723 CGCCCGCGCGCGCGCGCGG 47741
seq_name: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA194.DAT:AA073674
seq_documentation_block:
ID AA073674 standard; DNA; 2168 BP.
XX
AC AA073674;
XX
DT 21-JUN-1995 (first entry)
XX
DE Mycinamicin IV hydroxylating protein gene sequence.
XX
KM Mycinamicin; hydroxylase; macrolide; antibiotic; Micromonospora;
KM Micromonospora griseorubida; ds.
XX
OS Micromonospora griseorubida AT11725CN3.
XX
FT Key Location/Qualifiers
FT CDS 252..1442
FT FT /*tag=a
FT FT /product= mycinamicin IV hydroxylating protein
XX
PN JF06253853-A.
XX
PD 13-SEP-1994.
XX
PF 09-MAR-1993; 93JP-0047638.
XX
PR 09-MAR-1993; 93JP-0047638.
XX
PA (ASAH ) ASAH KASEI KOGYO KK.

```

```

XX WPI: 1994-338997/41.
DR P-PSDB; AAR60777.
XX DNA encoding a protein having mycinamicin IV hydroxylating
PT activity - for prodn. of mycinamicin, a macrolide antibiotic
XX
PS Claim 2; Page 12-14; 23pp: Japanese.
XX
CC The DNA sequence of a gene encoding a protein having mycinamicin IV
CC hydroxylase (MH) activity. The gene encodes a protein of 396 a.a. The
CC DNA was obtained from the macrolide antibiotic-producing bacterium
CC Micromonospora griseorubida AHA11725CN3/PTYS507. The gene was isolated
CC from the plasmid pTYS507. The protein encoded by this plasmid can be
CC used to produce mycinamicin IV in PTYS507-deficient Micromonospora
CC strains.
XX
SQ Sequence 2168 BP; 331 A; 704 C; 775 G; 358 T; 0 other;

alignment_scores:
Quality: 471.50 Length: 412
Ratio: 1.917 Gaps: 11
Percent Similarity: 59.709 Percent Identity: 32.039

alignment_block:
US-09-724-797-36 x AA073674 ..
Align seg 1/1 to: AA073674 from: 1 to: 2168
12 ThrAspAlaAspValArgArgAspProTyrProSerTyrHisThrLeu 28
|||||
252 ACTTCAGCTGAACCTAGGCGCTATCCATCAACAGTGCAGGCTTAC 301
28 u.....ArgHisAspProValHisArg. 35
302 CCTCGCTGCCCGGTACGGGAGCTGCAGAGACCGACCGGTCACCGGG 351
36 .....GlyAlaHisArgValTyrTyrValSerArgPheAlaAsp 48
|||||
352 TCCGGCGCGCTACAGGGGAGAGCGCTGCTGCTACCGCGTACGAGAC 401
49 ValArgAlaValLeuGlyAspGlnuArgPheAlaArgThrGlyTLeArgAr 65
|||||
402 GTCCGGCGCGTGTCTGCTGACGAGCACGCTTGTCCG. .... 437
65 gPheTrpThrAspLeuValGlyProGlyLeuLeu..... 76
438 .....GGCGCTGATGACCCCGGACGAGCGCTCGGA 468
77 .....AlaGluLeuValGlyAspPheLeuPheGlnAspGluProAsp 91
|||||
469 CCGGTCCCGGATGTCAGAGGGGCGCTGCTGCTGATGATCCGCCGAG 518
92 HisGlyArgLeuArgGlyValValGlyProAlaPheSerProSerAlaLe 108
|||||
519 CACAGCGGATGTCGCGGCTGCTGCAAGGCTTACCGCCGCGGCGC 568
108 uArgArgLeuGluProValIleAlaGlyThrValAspAspLeuArgP 125
|||||
619 AGATGGCGGCGACCGGCGAGCCAGCGACTGTGTCATGTTCGCGCCG 668
141 ProLeuAlaLeuArgAlaValLeuGlyLeuLeuGlyLeuProAlaAlaAs 157
|||||
669 CAGTTCGCGGTCAAGGATCTGCGAAGTCTGCTGCTCCGCGCGCGGA 718
157 TrpPGLyAlaValGlyArgTrpSerArgAspValGlyArgThrLeuAsp 174
|||||
719 CCACGACCGCTTACCGCGGTGCTCGGGGCTTCTCTCCACCGCGCGAG. 767

```

```

174 rgglValaserlaaglualuaspmetargArgglYHIsAlaAlaAlaIleaglu 190
768 .....GTACGGCGGAGGAGATGCAG.....GAGCGCGCGAGAGGCC 806
191 pheAlaasPTyrValgluArgAlaLeuAlaARgARgARgAGluGlyl 207
807 TACGCC...TACATGGCGACCTCATGCAGCCGAGCTCGAGAGAACCCAC 853
207 ygluaspLeuAlaLeuMetLeuAspAlaHisasp...ArgglyLeuM 223
854 GCAGCATCTGTGAGCGCGCTGTCCAGCCCGTGCACACAGAGACTGCT 903
223 eTserArgAsnGluIleValSerThrValAlaThrPheIlePhehrgly 239
904 TGTCCGACAGAGAACTGCTGCACCTTGCCATTGGCTGTCTGTCTGTGCGGA 953
240 HIsGluThrValAlaSerGlnValglYasnaAlaValleuSerLeuAl 256
954 TACGAGACAGACAGACCCAGATCGCGACTTCGTATCTCTGATGAC 1003
256 aHisProaspGlnLeuaspLeuAlaArgArgProaspLeuAlaag 273
1004 CCGACCGGAGCTGCGCTGCAGACACTGATCGATCGCGGAGCTGATCCGT 1053
273 InAlaValgluGluCysLeuArgTyrAspProSerValglSerAsnThr 289
1054 CCGAGTCGAGAGAGCTGACGGGTGGCGCGCGCGGGGTGGGAAGCGG 1103
290 .....ArgGlnLeuaspValaspValgluLeuArgglYArgVal 304
1104 TTCCCGCGCTACGCGGTGGAGATGTCACGTTCGCGGGCGCGACCATCCG 1153
304 gArgaspaspValValValAlaValleuAlaValAlaAsnArgaspPro 321
1154 CGCGGGCGAGCGCGGTGCTCGCTGCAGCGGTGCCGACCGGAGAGCAGG 1203
321 rgaTyrTyrAspArgProaspAspPheaspIleGluArgaspProVal 337
1204 CCCAGTTTCCCGACGCGGACCGGATGCAGTCGACCGACCCCAATCAG 1253
338 SerMetSerPheglYalaglyMetArgTyrCysLeuGlySerTyrLeu 354
1254 CATCTCGGTTCGGACACGCGGTCCACCATGCTCGCGGTCTCTCTGCG 1303
354 aArgThrGlnLeuArgAlaAlaVal...AlaAlaLeuAlaArgLeuPro 370
1304 CCGGGTGGAACTCCAGGTGCAGTCGAGGTCTCTCCAGGACTGCCAG 1353
370 lYleuAlaGluGlyCysAlaSerAspAlaLeuAlaTyrGlnProArgThr 386
1354 GGATTCGCGCTGGGATTCGCCGAGACGACGCTGCGGTGAGGAGAGGATG 1403
387 MetPheArgglYleuAlaSerLeuProIleAlaPhe 398
1404 TTGCTGCTGTGCGCCACTCGAGCTCCGCGTGTGG 1439

seq_name: /sids1/gc9data/geneseq/geneseq-emb1/NA2001A.DAT.AA171296
seq_documentation_block:
ID AA171296 standard; DNA; 2795 BP.
XX
AC AA171296;
XX
DT 24-JAN-2002 (first entry)
XX
DE Streptomyces griseus RppA protein encoding DNA SEQ ID NO.1.
XX
KW Streptomyces griseus; RppA; pyran-2-one; drug; agricultural; ds.
XX
OS Streptomyces griseus.
XX
FH Key Location/Qualifiers

```

```

FT CDS 1596..2714
FT /tag=a
FT /product="RppA protein"
PN JP2001238677-A.
PD 04-SEP-2001.
XX
XX 01-MAR-2000; 2000JP-0055782.
XX
XX 01-MAR-2000; 2000JP-0055782.
XX
XX (MTR ) MITSUBISHI RAYON CO LTD.
XX
XX WPI: 2001-613891/71.
XX
XX P-PSDB; AAG68148.
XX
XX Preparation of pyran-2-one derivative, used as intermediate and as
XX additive for thermosensible transfer paper, comprises contacting
XX cardonyl compound with microbe containing polyketide synthase gene -
XX
XX Claim 6; Page 7-8; 11pp; Japanese.
XX
XX The present invention describes the preparation of pyran-2-one
XX derivatives. The preparation method comprises contacting carbonyl
XX compounds with microbes containing a polyketide synthase gene.
XX Pyran-2-one derivatives can be used as synthetic intermediates for
XX drugs and agricultural chemicals. They can also be used as additives
XX for thermosensible transfer paper. The present sequence encodes the
XX specifically claimed Streptomyces griseus RppA protein, which is used
XX in the exemplification of the present invention.
XX
XX Sequence 2795 BP; 435 A; 1051 C; 908 G; 401 T; 0 other;

```

```

alignment_scores:
Quality: 461.00 Length: 412
Ratio: 1.837 Gaps: 15
Percent Similarity: 60.922 Percent Identity: 33.495

```

alignment_block:

US-09-724-797-36 x AA171296 ..

Align seg 1/1 to: AA171296 from: 1 to: 2795

```

16 ValArgArgAspProTyr.....Prose 23
374 GTCCGAGAACTGTCCTTCGACTACGCGCAGAGCTGAGTGCAGCCCA 423
23 rTyrHISrPLeuLeuArgHISAspProValHISArg..... 35
424 GCTCAGGCAATTCCTACCGAGGAGCGGTGTCGCCATCGTAGCGCT 473
36 ..GlyAlaHisArgValTyrTyrValSerArgPheAlaaspValArgAla 51
474 ACGAGAGAGGCGAGCGCTGCTGTCACCCGCTAGAGAGAGCTCGGAGC 523
52 ValLeuGlyAspGluArgPheAlaArgThrGlyIle.....ArgArgPh 66
524 GTCACACCGACCGCGGCTTCAGCGCAGCGCTCTCGCGCGGAGACTT 573
66 eTrrPThrAspLeuValglYProglYleuLeuAlaGlu.....IleValG 81
574 C.....CCCCGATGACCCCGAGCGGATCGTGCAGG 605
81 lYAspIleIleLeuPheGlnAspGluProAspHisglYArgLeuArggly 97
606 CCGAGTCATCAACCTCATGAGACCGCGGACGAGCGGCTGCGGAGG 655
98 ValValglYProAlaPheSerProSerAlaLeuAlaGluGluProVa 114
656 CTGGTGGCCAAAGAGCTTCACCCG.....CGTCGTCGAGACAGAT 696

```



```

114 111aaaglythr.....ValAspAspIleuArgProAlaLeuA 128
      : : : : : : : : : : : : : : : : : : : : : : : :
697 GCGCGCGGAGCCAGCCGCTGTGTGACCGCTGTGTGACGAGATGAGG 746
      : : : : : : : : : : : : : : : : : : : : : : : :
128 IaArgGlyAlaMet...AspValAlaSpGlyLeuAlaTyrProLeuAla 143
      : : : : : : : : : : : : : : : : : : : : : : : :
747 AGAGAGCTCAACGGGGGACATTGCTGCGCGGGGTCTCCGGCGCTGGCG 796
      : : : : : : : : : : : : : : : : : : : : : : : :
144 LeuArgAlaValLeuGlyLeuLeuGlyLeuProAlaAlaAspTrpGlyAl 160
      : : : : : : : : : : : : : : : : : : : : : : : :
797 CTGATACACCATCTGGCAGAGGACATCCACATCCCGAGGCGGAGC..... 838
      : : : : : : : : : : : : : : : : : : : : : : : :
160 aValGlyArgTrpSerArgAspValGlyArgTrpLeuAspArg.....G 175
      : : : : : : : : : : : : : : : : : : : : : : : :
839 ....CGCGCCCTGTGCTCCGGGGCCAGGCCCATGACATGATGAGACGTGGGG 884
      : : : : : : : : : : : : : : : : : : : : : : : :
175 IyAlaSerAlaGlyAspMetArgArgGlyNysAlaAlaIleAlaGluPro 191
      : : : : : : : : : : : : : : : : : : : : : : : :
885 CCGCGGGGACAGCAGGAGACGGGTGGCGC.....GCCAAGGCGGAGCTG 925
      : : : : : : : : : : : : : : : : : : : : : : : :
192 AlaAspTyrValGlyArgAlaLeuAlaArgArgArgGlyGlyGly 208
      : : : : : : : : : : : : : : : : : : : : : : : :
926 CGCGGCTACTTCCAGAGAGCTGACCGCGGACCGCGCGCTCCCGGGCGGA 975
      : : : : : : : : : : : : : : : : : : : : : : : :
208 uAspIleuAlaLeuMetLeuAspAlaIleAspArgGly...LeuMetS 224
      : : : : : : : : : : : : : : : : : : : : : : : :
976 GGACTCTATCAGCACACCTGTGGCCACCGCGGGGACGGCAGACACTGTGG 1025
      : : : : : : : : : : : : : : : : : : : : : : : :
224 eArgAsnGluIleValSerThrValValThrPheIlePheTrpGlyNis 240
      : : : : : : : : : : : : : : : : : : : : : : : :
1026 ACGAGACGACAGCTGGCGGTGATGCGCATGCGCTGTCTCATCACCGGCGAG 1075
      : : : : : : : : : : : : : : : : : : : : : : : :
241 GluThrValAlaSerGlnValGlyAsnAlaValLeuSerLeuLeuAlaH 257
      : : : : : : : : : : : : : : : : : : : : : : : :
1076 GACACACGACAGCTTACAGCTGTGGCAACATCGCTACACCTGTCTACCGG 1125
      : : : : : : : : : : : : : : : : : : : : : : : :
257 sProAspGlnLeuAspLeuAlaTyrArgArgProAspLeuLeuAlaGlnA 274
      : : : : : : : : : : : : : : : : : : : : : : : :
1126 CCGGAGACCTGTGGGTCTCCCTCCGGGCGCAACCGCAGCGGTGCTCCCGCA 1175
      : : : : : : : : : : : : : : : : : : : : : : : :
274 IaValGluGluCysLeuArgTyrAspPro.....SerValGlnSer 287
      : : : : : : : : : : : : : : : : : : : : : : : :
1176 CCGTGTGAGAGAGCTGTGCGCCACATCCCTCCCGCAAGGGCGTGGCATC 1225
      : : : : : : : : : : : : : : : : : : : : : : : :
288 AsnThrArgGlnLeuAspValAlaGluLeuArgGlyArgArgLeuArg 304
      : : : : : : : : : : : : : : : : : : : : : : : :
1226 CCGCGATGCGCCTGTGAG...GACGTGAGAGCTGTCCGGCTGCTCATCAAA 1272
      : : : : : : : : : : : : : : : : : : : : : : : :
304 gArgAspAspValValValValLeuAlaGlyAlaIleAsnArgAspPro 321
      : : : : : : : : : : : : : : : : : : : : : : : :
1273 GCGCGCGGACGTGTGCTACGTGTCTCACTGACGCGGCCAACCGGAGATCCG 1322
      : : : : : : : : : : : : : : : : : : : : : : : :
321 rGArgTyrAspArgProAspAspPheAspIleGluArgAspProValPro 337
      : : : : : : : : : : : : : : : : : : : : : : : :
1323 CCAAGTTCACCGCTCCCGAGACCTGGACCCGACGCGGACACATCCCC 1372
      : : : : : : : : : : : : : : : : : : : : : : : :
338 SerMetSerPheGlyAlaGlyMetArgTyrCysLeuGlySerTyrLeuAl 354
      : : : : : : : : : : : : : : : : : : : : : : : :
1373 CACATGACCTTGTGGCTGGCGCGCCACACCATGCTGTGGCGCGCGCTGGC 1422
      : : : : : : : : : : : : : : : : : : : : : : : :
354 aArgTrpGlnLeuArgAlaAlaValAlaAla...LeuAlaArgLeuProG 370
      : : : : : : : : : : : : : : : : : : : : : : : :
1423 CACACATGAGAGCTGGAAGTGGCTTTCACAGCTGCTACCCGCTCCCGG 1472
      : : : : : : : : : : : : : : : : : : : : : : : :
370 IyLeuArgLeuGlyCysAlaSerAspAlaLeuAlaTyrGlnProArgThr 386
      : : : : : : : : : : : : : : : : : : : : : : : :
1473 CCTGTGCTGTGACGTGCGCGCCGAGAGACGTCTGCTGGAACACAGACGTCC 1522
      : : : : : : : : : : : : : : : : : : : : : : : :
387 MetPheArgGlyLeuAlaSerLeuProIleAlaPhe 398
      : : : : : : : : : : : : : : : : : : : : : : : :
1523 ATCTTGGCGTTTACCCTGCGCTGCGCCGTGACATGG 1558

```

```

seq_documentation_block:
ID   AAF81954 standard; DNA; 2795 BP.
XX
XX   AAF81954;
XX
XX   18-JUN-2001 (first entry)
XX
XX   Streptomyces griseus actinomycetes polyketide synthase encoding DNA.
DE
XX   Streptomyces griseus; actinomycete; polyketide synthase; rppA; drug;
KW   1,3,6,8-tetrahydroxynaphthalene; melanine; agricultural chemical;
KW   malonyl CoA; ds.
XX
XX   Streptomyces griseus.
OS
XX
XX   Key Location/Qualifiers
FH   CDS 1596..2714
FT   /tag= a
FT   /product= "Actinomycetes polyketide synthase"
FT
XX
XX   JP2000342269-A.
XX
XX   12-DEC-2000.
XX
XX   02-JUN-1999; 99JP-0155526.
XX
XX   02-JUN-1999; 99JP-0155526.
XX
XX   (MTR ) MITSUBISHI RAYON CO LTD.
XX
XX   WPI; 2001-303158/32.
XX   P-PSDB; AAB74843.
XX
XX   Preparation of 1,3,6,8-tetrahydroxynaphthalene for use as a synthetic
XX   material for melanine and as an intermediate for drugs and agricultural
XX   chemicals, comprises using a microbe and an enzyme
XX
XX   Claim 2; Page 6-7; 10pp; Japanese.
XX
XX   The present invention describes a method for preparing
XX   1,3,6,8-tetrahydroxynaphthalene comprising using a microbe cell
XX   containing an rppA gene encoding actinomycetes polyketide synthase, to
XX   convert malonyl CoA as the substrate or a substance which can produce
XX   malonyl CoA in the microbe cell to 1,3,6,8-tetrahydroxynaphthalene and
XX   recovering it. 1,3,6,8-tetrahydroxynaphthalene is used as a synthetic
XX   material for melanine and as an intermediate for drugs and agricultural
XX   chemicals. The present sequence encodes the actinomycetes polyketide
XX   synthase protein isolated from Streptomyces griseus.
XX
XX   Sequence 2795 BP; 435 A; 1051 C; 908 G; 401 T; 0 other;
XX
XX   alignment_scores:
XX       Quality: 461.00 Length: 412
XX       Ratio: 1.837 Gaps: 15
XX   Percent Similarity: 60.922 Percent Identity: 33.495
XX
XX   alignment_block:
XX   US-09-724-797-36 x AAF81954 ..
XX
XX   Align seq 1/1 to: AAF81954 from: 1 to: 2795
XX
XX   16 ValaTgATgAaSPProTyr.....Prose 23
XX   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
374 GTCCGGAACTGCTCTTTCGACTACGCCGACGAGCTGAGTTCGACCCCA 423
XX
XX   23 rTynHtStRDeuLarHgnHsAProValHnATg..... 35
XX   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
424 GTCGAGCAATGCTGCACCGAGGACGCGGTGCCCGCATCGGTGCGGT 473
XX
XX   36 ..GtYAlaHnATgValTTrPTyValSerArGrPhnElaaSPvalATgAla 51
XX   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```


Ratio: 1.814 Gaps: 13
Percent Similarity: 59.390 Percent Identity: 30.282

alignment_block:

US-09-724-797-36 x AA014548 ..

Align seg 1/1 to: AA014548 from: 1 to: 1400

```

11 ProthraspaValaArgArgspProtyrPro..... 22
13 CTACCCGACGATGAAAGTGTATGACACCGTTCGATTCGAAAGCGA 186
23 .setyrhis.....Trrleu.....leuarghis.... 30
187 CTCCTTCACGTCGACGTGTACCGACCTACGCCGACGCTGCCGAGACG 236
31 .....AspProValHisArgGlyAlaHisArgValTrrTyrrValSer 44
237 CCGCGGTGACCGCGGTGCTTCCTCGGGCAGGACGCTGGCTGATCAC 286
45 ArgPheAlaAspValArgAlaValleuGlyAspGluArgPheAlaArgTh 61
287 GGCTACGACGAGCGAAGCGCGCTGAGCAGCTGCGCTGAGCAGCGA 336
61 rglYlIleArgArgPheTrrPThrAspLeuValGlyProGlyLeuAlaG 78
337 CCGGAAGAAGATGAC.....CCGGCGCTGGAGGTGCG 368
78 IuIleValGly.....AspIleIleLeuPhe..... 86
369 AGTTCGCGCATACCTCGGTTTCCCGGAGGACGTGCGAATCTACTTCGC 418
87 .....GlnAspGluProAspHisGlyArgLeuArgGlyVal 98
419 ACCAACATGGGACACGAGCGACCGCGACCCACACCGCGCTGCGCACT 468
98 lValGlyProAlaPheSerProSerAlaLeuArgArgLeuProValI 115
469 GGTGTGCGAGAGTTCACCGTCCGCGGTGAGCGCATGGGCGCCCGCG 518
115 lValGlyThrValAspAspLeuLeuArgProAlaLeuAlaArgGlyAla 131
519 TCGACGACATCCCGCGAGCTGTCGACGAGTGGGCGACTCGCGGCTG 568
132 MetAspValAlaAspGluLeuAlaTyrProLeuAlaLeuAlaValle 148
569 GTCGACATCGTGTACCGCTTCGCCACCCGCTGCCCATCAAGTCATGT 618
148 uGlyLeuLeuGlyLeuProAlaAlaAspTrrPglyAlaValGlyArgTr 165
619 CGAGCTGCTCGGCTGCGACGAAATACCGCGGAGATTCGGGCGGTGA 668
165 eArgAspValGlyArgThrIleuAspArgGlyAlaSerAlaGluAspMet 181
669 GCTCGGAGATC...CTGTCATGACCCGCGAGCGGGCGAAACAG.... 709
182 ArgArgGlyHisAlaAlaIleAlaGluPheAlaAspTyrValGluArgAl 198
710 ...CGCGGCGAGGCGGCC...AGGAGGTGCTGCAATTCATCTTCGACT 753
198 AlauAlaArgArgArgGlyGlyGlyAlaAspLeuAlaLeuMetL 215
754 GGTGAGGCGCGCGCACCGAGCCGCGGAGACCTGTCGCGGCTTA 803
215 euAspAlaHis.....AspArgGlyLeuMetSerArgAspGluIleVal 229
804 TCAGGGTCCAGGACGACGATGACGCTGCTGACGCGCCGAGAGTGTTC 853
230 SerThrValAlaThrPheIlePheThrGlyHisGluThrValAlaSerG 246
854 TCATCGCGCGCTGTGCTGTGCTGCGCGGTTCGAGGCGGTGAGGCT 903
246 nValGlyAsnAlaValLeuSerLeuAlaHisProAspIleuAspRL 263

```

```

904 CATGGGATGGGACACTACTGCTGCTACCCACCGCGGACCTGCGCGC 953
263 euleuArgArgArgProAspLeuAlaGlnAlaValGluGlyCysLeu 279
954 TGTGTGCGGCGGAGCCGTGCGCGCTGCGCAACCGCTGAGAGATCTG 1003
280 ArgTyrAspProSerValGlnSerAsnThrArgGlnLeuAspValAsp 296
1004 CGTACATCGCTGTGACGAGACACACCGCGCTGCGCGGAGAGACT 1053
296 lGluLeuArgGlyArgArgLeuArgArgAspAspAlaValAlaLeu 313
1054 GGAATGCGGCGGTGCTGCGCATCCCGCATACACGAGCTGTGTCGCA 1103
313 lArgIAlaAlaAsnArgAspProArgArgTyrArgAspArgProAspPhe 329
1104 ACGCGCGGCGCAACCGCGAGCCGAAAGCATTTCCGCGACCCGCTCC 1153
330 AspIleGluArgAspProValProSerMetSerPheGlyAlaGlyMetAr 346
1154 GACGTCAACCGCGACACCGCGGCCACCTGTGTCGGCGAGGCAATCCA 1203
346 gTyrCysLeuGlySerTyrLeuAlaArgThrGlnLeuArgAlaAlaValA 363
1204 CTTCGTGATGGCGCGCGCTGCGCAAGCTGAGGCGAGAGTGGCTGCG 1253
363 lAlaLeu...AlaArgLeuProGlyLeuArgLeuGlyCysAlaSerAsp 378
1254 GCGCGCTGTGCGCGCGCTGCCGCTGCTGCTGCGGAAATCGACGCGGAC 1303
379 AlaleuAlaTyrGlnProArgThrMetPheArgGlyLeuAlaSerLeuP 395
1304 GACGTGCTGTGCGCGCTGCGGTGCTGCTGCGGAGATGACCACTTACC 1353
395 oIleAlaPhe.....ThrProGly 401
1354 GGTGCGCTGCGAGCATGACACACTGGC 1381

seq_name: /STD1/gcgsdata/geneseq/geneseq-emb1/NA2000.DAT:AAA9567
seq_documentation_block:
ID AAA95667 standard; DNA; 1204 BP.
XX
AC AAA95667;
XX
DT 14-FEB-2001 (first entry)
XX
DE Bacillus subtilis hydroxylating enzyme gene #2.
XX
KW Hydroxylated bicyclic compound; lactone; HMG-CoA; inhibitor; ds;
KW hydroxymethylglutaryl co-enzyme A reductase; hypocholesterolaemia.
XX
OS Bacillus subtilis.
XX
PN WO20044886-A1.
XX
PD 03-AUG-2000.
XX
PF 28-JAN-2000; 2000WO-JP00472.
XX
PR 29-JAN-1999; 99JP-0021707.
XX
PA (KYO) KYOMA HAKKO KOGYO KK.
XX
PI Endo H, Yonetani Y, Mizoguchi H, Hashimoto S, Ozaki A;
XX
DR MPI: 2000-548827/50.
XX
DR P-PSDB; AAB15502.
XX
PT New protein derived from Bacillus genus microorganism useful for
PT producing a hydroxylated bicyclic compound e.g. HMG-CoA reductase
PT inhibitors

```

PS Claim 16; Page 90-94; 11pp; Japanese.
XX
CC This sequence represents the coding region for a novel *Bacillus* derived
CC protein having the activity of producing a hydroxylated bicyclic compound
CC or the corresponding lactone from a bicyclic compound or the
CC corresponding lactone. The protein is used for preparing reduced
CC compounds useful as hydroxymethylglutaryl co-enzyme A (HMG-CoA) reductase
CC inhibitors and hypocholesterolaemic agents.
XX
SQ Sequence 1204 BP; 333 A; 253 C; 316 G; 302 T; 0 other;

alignment_scores:
Quality: 449.50 Length: 358
Ratio: 2.007 Gaps: 7
Percent Similarity: 62.570 Percent Identity: 30.447

alignment_block:
US-09-724-797-36 x AAA95667 ..

Align seq 1/1 to: AAA95667 from: 1 to: 1204

20 ProTyrProSerTyrHisThrLeuArgHisAspProVal...HisAr 35
|||:||||| |||:| :|:| :|:| |||||
83 CCGTTCATCGTATGATCGATCGAAGAAAGATGCCCTGTTCTTGA 132
|||:||||| |||:| :|:| :|:| |||||
35 gGlyAlaHisArgValTyrTyrValSerArgPheAlaaspValArgalav 52
|||:||||| |||:| :|:| :|:| |||||
133 TGAAGAAACCAAGTGTGAGCGTCTTCTTATGATGATGTCAAAAG 182
|||:||||| |||:| :|:| :|:| |||||
52 alLeuGlyAspGluArgPheAlaArgThrGlyIleArgArgPheTyrThr 68
|||:||||| |||:| :|:| :|:| |||||
183 TTGTGGGATTAAGAGTGTCTTCCAGTGTGATGCCGAG..... 223
|||:||||| |||:| :|:| :|:| |||||
69 AspLeuValGlyProGlyLeuLeuAlaGluIleValGlyAspIleIleLe 85
|||:||||| |||:| :|:| :|:| |||||
224CAGCAAGCTCTATTGGAAATCCATCAT 252
|||:||||| |||:| :|:| :|:| |||||
85 upheGlnAspGluProAspHisGlyArgLeuArgGlyValAlaGlyProA 102
|||:||||| |||:| :|:| :|:| |||||
253 TAAACATGACCCGCCAGCATACAAAATTCGTTCACTGTAACAAG 302
|||:||||| |||:| :|:| :|:| |||||
102 laPheSerProSerAlaLeuArgArgLeuGluProValIleAlaGlyThr 118
|||:||||| |||:| :|:| :|:| |||||
303 CTTTACTCCGCCGATGAAGCAATGGGAACCGAATTCAGAAATC 352
|||:||||| |||:| :|:| :|:| |||||
119 ValAspAspLeuArgProAlaLeuAlaArgGlyAlaMetAspValVa 135
|||:||||| |||:| :|:| :|:| |||||
353 ACAAGTGAATGATTCAAAAATTTCAAGGGCGCAGTGTGACCTTGT 402
|||:||||| |||:| :|:| :|:| |||||
135 laSpGluLeuAlaTyrProLeuAlaLeuArgAlaValLeuGlyLeuLeu 152
|||:||||| |||:| :|:| :|:| |||||
403 TCACGATTTTTCATCCCGCTCCGTTATGTGATATCGAGCTGCG 452
|||:||||| |||:| :|:| :|:| |||||
152 lyeuProAlaIleAspTyrGlyAlaValAlaGlyTyrSerArgAsp... 167
|||:||||| |||:| :|:| :|:| |||||
453 GAGTGCCTTACGCGCATATGGAACAGCTTAAAGCATGTCT...GATCTT 499
|||:||||| |||:| :|:| :|:| |||||
168 ...ValGlyArgThrLeuAspArgGlyAlaSerAlaGluAspMetArgAr 183
|||:||||| |||:| :|:| :|:| |||||
500 CTGGTCAGTACACGAGAGATAAAGTAGAGAACTGAA..... 538
|||:||||| |||:| :|:| :|:| |||||
183 gGlyHisAlaIleAlaGluPheAlaAspTyrValGluArgAlaLeuA 200
|||:||||| |||:| :|:| :|:| |||||
539AAAGCTTTTGGAAAGACGAGATTAAGTAGAGAGAACTGG 581
|||:||||| |||:| :|:| :|:| |||||
200 la.....ArgArgArgArgGlyGlyGlu 208
|||:||||| |||:| :|:| :|:| |||||
582 CCGCGTTTTCGCGCATCATAGAAAGAAACGAAACAAACGGAACAG 631
|||:||||| |||:| :|:| :|:| |||||
209 AspLeuLeuAlaLeuMetLeuAspAlaHisArgGly...LeuMetSe 224
|||:||||| |||:| :|:| :|:| |||||

632 GATATATTCTATTATTAGTGAAGCGGAAGAAACAGCGACACTGTC 681
224 rArgAsnGluIleValSerThrValValThrPheIlePheThrGlyHisG 241
|||:||||| |||:| :|:| :|:| |||||
682 CGGTGAAGAGCTGATCCGTTGTGCACGCTGCTGGTGGCGGGAATG 731
241 lufhValAlaSerGlnValGlyAsnAlaValLeuSerLeuLeuAlaHis 257
|||:||||| |||:| :|:| :|:| |||||
732 AAACCACTACCAACCTGATTTCAAATGCCATGTACAGCATATTAAACG 781
258 ProAspGlnLeuAspLeuLeuArgArgArgProAspLeuAlaGlnAl 274
|||:||||| |||:| :|:| :|:| |||||
782 CCAGCGCTTTACGAGAACTGCCACACCATCTGAACTGATCCACAGC 831
274 aValGluGluCysLeuArgTyrAspProSerValGlnSerAsnThrArg 291
|||:||||| |||:| :|:| :|:| |||||
832 AGTGAAGAAACCTTGGCTTTCAGACGCCGCCGCTTTGAGAGCCA 881
291 lneuAspValAspValGluLeuArgGlyArgArgLeuArgArgAspAsp 307
|||:||||| |||:| :|:| :|:| |||||
882 TTGCCAAGCGGATACGAGATCGGGGGCACCTGATTAAAGAACTGAT 931
308 ValValValValLeuAlaGlyAlaAlaAsnArgAspProArgArgTyrAs 324
|||:||||| |||:| :|:| :|:| |||||
932 ATGCTTTTGGCGTTTGGCATCGCAATCGATGAAAGAACTTTGA 981
324 pArgProAspAspPheAspIleGluArgAspProValProSerMetSerP 341
|||:||||| |||:| :|:| :|:| |||||
982 CAGACCCACCATGTTGATATCCGCCGCATCCCAATCCGATATATGCGT 1031
341 heGlyAlaGlyMetArgTyrCysLeuGlySerTyrLeuAlaArgThrGln 357
|||:||||| |||:| :|:| :|:| |||||
1032 TTGGCCACGCGATCATTTTGGCTTGGGGCCCGCTTCCGCTTGAA 1081
358 LeuArgAlaAlaValAlaAlaLeu 365
|||:||||| |||:| :|:| :|:| |||||
1082 GCAATATTCGCTTAACGCTCTTG 1105

...

alignment_block:
US-09-724-797-36 x US-09-407-549-1 ..

Align seg 1/1 to: US-09-407-549-1 from: 1 to: 8478

```

8 AlaphaspProThrAspAlaaspValArgargProTyrProSerTy 24
  ||| : : : : : : : : : : : : : : : : : : : : : : : :
5493 GCATCGTCACACTGCTCTGAGTTTGGAAACCATATCTCTTTT 5542
24 rHistrpleuLeuArgHisaspProValHisArgGlyAla.....H 38
  ||| : : : : : : : : : : : : : : : : : : : : : : : :
5543 GCACACATTGCGAGCTGTTCATCTCATTAAGAGGAGTTTCTTAAT 5592
38 iArgValTryptValSerArgPheAlaaspValArgAlaValLeuGly 54
  : : : : : : : : : : : : : : : : : : : : : : : :
5593 ACCCGGCTGATGTACAGGATATGAGAAACGGCTGCTATTTTGA 5642
55 AspGluArgPhe...AlaArgThrGlyLeuArgArgPheTrrPhe 70
  ||| ||||| : : : : : : : : : : : : : : : : : :
5643 GATGGAGATTCAAAGTCCGACCCCGCTGAGAGCTCAACCAATA 5692
70 uValGlyProGlyLeuLeuAlaGluValValGlyAspIleIleuPhe 87
  ||| : : : : : : : : : : : : : : : : : : : : : : : :
5693 TCAGGAC.....CTTCACATGTGCAAAATCAATGATGCTGTTTC 5733
87 InaspGluProAspHisGlyArgLeuArgGlyValAlaGlyProAlaPhe 103
  ||| : : : : : : : : : : : : : : : : : : : : : : : :
5734 AGAACACCGCTGATCATATGACGATTTGCGAGCGTTCGCGAGCGTT 5783
104 SerProSerAlaLeuArgArgLeuGluProValIleAlaGlyThrValas 120
  : : : : : : : : : : : : : : : : : : : : : : : :
5784 ACGCCGAGACGACAGAGATTATCAGCCGTATCATTTGAAGACTGTCCA 5833
120 paspleuLeuArgProAlaLeuAlaArgGlyAlaMetAspValaspg 137
  ||||| : : : : : : : : : : : : : : : : : : : : : : : :
5834 TCATTGCTTATCATAGTCAGCAAGGTAAGAAAAAAGATGGAGTTCATTCGG 5883
137 IuLeuAlaTyrProLeuAlaLeuArgAlaValLeuGlyLeuGlyLeu 153
  : : : : : : : : : : : : : : : : : : : : : : : :
5884 ACTTGCTTTTCCTTTAGCAAGTTTTCATATGCAACATATATAGGTGA 5933
154 ProAlaAlaAspTrrPglyAlaValGlyArgTrrSerArgAspValGlyArg 170
  ||| : : : : : : : : : : : : : : : : : : : : : : : :
5934 CCGGAGGAGATAGGAGCAATTAAGAGAGTGGCTGCGAATCTCATTTCA 5983
170 gThrLeuAspArgGlyAlaSerAlaGluAspMetArgArgGlyHisAla 187
  : : : : : : : : : : : : : : : : : : : : : : : :
5984 AACGATTATTTTACCGGCTCAAGAAAGGCTTTAACAGAGGCAATATTA 6033
187 IalIleAlaGluPheAlaAspTrrValGluArgAlaLeuArgArgArg 203
  : : : : : : : : : : : : : : : : : : : : : : : :
6034 TGGCTGTGACAGCTATGCGCATATTTCAAGAGCTGATTCAAAGAGAAA 6083
204 ArgGluGlyGlyGluAspLeuLeuAlaLeuMetLeuAspAlaHisAspArg 220
  ||||| : : : : : : : : : : : : : : : : : : : : : : : :
6084 CGCCACCCCTCAACAGATATGATCAGCATGCTCTTGAAGGGGAGAGAAA 6133
220 gGlyLeuMetSerArgAsnGluIleValSerThrValValThrPheIleP 237
  : : : : : : : : : : : : : : : : : : : : : : : :
6134 GGATTAAGTGCAGAGAGAGAGGCGGCATCTACGTCATATTTGCTGGCGA 6183
237 heHrGlyHisGlyThrValAlaSerGluValGlyAsnAlaValLeuSer 253
  : : : : : : : : : : : : : : : : : : : : : : : :
6184 TCGCCGCGCATAGAGCAACGGTCAATCTCATACCAATTCAGTCTTTGT 6233
254 LeuLeuAlaHisProAspGluLeuAspLeuLeuArgArgArgProAspLe 270
  ||||| ||||| : : : : : : : : : : : : : : : : : :
6234 CTGCTGCAGCATCCAGAACGCTTTTAAATCAGAGAAATCCAGATCT 6283
270 uLeuAlaGluAlaValGluGlyCysLeuArgTyrAspProSerValGlns 287
  : : : : : : : : : : : : : : : : : : : : : : : :
6284 TATTGGTAACCGCAGTGCAGAGATGTTTACGCTATGAAGCCCAACGAAA 6333

```

```

287 eRasNThrArgGluLeuAspValaspValIuLeuArgGlyArgArgLeu 303
  : : : : : : : : : : : : : : : : : : : : : : : :
6334 TGACAGCCAGAGTTGCTGACAGAGATATGACATCTGCGGGGACGATC 6383
304 ArgArgAspAspValValValLeuAlaGlyAlaAlaAsnArgAspTr 320
  ||| : : : : : : : : : : : : : : : : : : : : : : : :
6384 CGTAAAGAGAACAGTCTATCTTTGTGTAGGAGCGGCTATATCGAAGCC 6433
320 oArgArgTyrAspArgProAspAspPheAspIleGluArgAspProValP 337
  : : : : : : : : : : : : : : : : : : : : : : : :
6434 TAGCATATTCACAGAACCCGATGCTTCGATATTTACAGAGAGCTTATC 6483
337 roSerMetSerPheGlyAlaGlyMetArgTyrCysLeuGlySerTyrLeu 353
  ||| : : : : : : : : : : : : : : : : : : : : : : : :
6484 CGCATCTTTATTCGCGCATCGGCATCATGTTTGGCTTAGGCTCTCGCTG 6533
354 AlaArgThrGluLeuArgAlaAlaVal...AlaAlaLeuAlaArgLeuP 369
  ||||| : : : : : : : : : : : : : : : : : : : : : : : :
6534 GCACGATTAGAACGCAATTCGATTAACACATCTTCTGCGAGGAATGCC 6583
369 oGlyLeuArgLeuGlyCysAlaSerAspAlaLeuAlaTyrGlnProArgT 386
  : : : : : : : : : : : : : : : : : : : : : : : :
6584 CAGCCTTAATCTT.....GCGGATTTTGAATGCGCGTATCGCGCCTTT 6627
386 hMetPheArgGlyLeuAlaSerLeuProIleAlaPhe 398
  ||||| : : : : : : : : : : : : : : : : : : : : : : : :
6628 TTGGATTTCGGGCGCTTGAGAGAGCTGCCGCTGACTTTT 6665
seq_name: /cgn2_6/ptodata/1/lna/6B_COMB.seq:US-09-105-537-38
seq_documentation_block:
; Sequence 38, Application US/09105537A
; Patent No. 6265202
; GENERAL INFORMATION:
; APPLICANT: Sherman, D. H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikomycin
; FILE REFERENCE: 600.4380U1
; CURRENT APPLICATION NUMBER: US/09/105,537A
; CURRENT FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO: 38
; LENGTH: 1251
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
; US-09-105-537-38
alignment_scores:
Quality: 538.50 Length: 398
Ratio: 2.216 Gaps: 8
Percent Similarity: 61.055 Percent Identity: 34.673
alignment_block:
US-09-724-797-36 x US-09-105-537-38 ..
Align seg 1/1 to: US-09-105-537-38 from: 1 to: 1251
15 AspValArgArgAspProTrrPheSerTrrHisTrrPleuLeuArgHisAs 31
  ||| : : : : : : : : : : : : : : : : : : : : : : : :
67 GATTTTGGCGGCGGATCCGATTCGAGCTACGCGAGACTGCGTCCGAGGG 116
31 pProValHisArg.....GlyAlaHisArgValTrrTrrValas 44
  ||||| ||||| : : : : : : : : : : : : : : : : : :
117 TCGGCGCCACCGGATCGCAGCCCGGAGGAGAGGAGTGTGGCTGTGCG 166
44 eArgArgPheAlaAspValArgAlaValLeuGlyAspGluArgPheAlaArg 60
  : : : : : : : : : : : : : : : : : : : : : : : :
167 TCGGCTACGACCGGCGGCGGCTTCGCGCATCCCGGTTTCAGCAAG 166

```



```

61 ThrGlyIleArgArgPheTrpThrAspLeuValGlyProGlyLeuLeuAl 77
   ||| : : : : : |||
217 GAC.....TGGCGCAACTCCACGACTCCCTGACCGAGGC 251
   ||| : : : : : |||
77 aglu...IleValGlyAspIleIleLeuPheGlnAspGluProAspHisG 93
   ||| : : : : : |||
252 CGAGCGCGCGCTCAACCAACATGCTGTGAGTCCGACCCCGCGGCACA 301
   ||| : : : : : |||
93 LysArgLeuArgGlyValValGlyProAlaPheSerProSerAlaLeuArg 109
   ||| : : : : : |||
302 CCGCGCTGCGCAAGCTGTGTGGCCGTGATCCATCGCCCGGTTGAG 351
   ||| : : : : : |||
110 ArgLeuGluProValIleAlaGlyThrValAspAspLeuArgProAl 126
   ||| : : : : : |||
352 TTGCTGGGGCGCCCGGCTCCAGAGATGTGTGACGGGCTCTGGAGCGCAT 401
   ||| : : : : : |||
126 aleuAlaArg.....GlyAlaMetAspValValAspGluLeuAlaTyrP 141
   ||| : : : : : |||
402 GCTGGCGCGCGCGCGCGCGCGCGCATGTGATGAGTCCCTGGCGCTGGC 451
   ||| : : : : : |||
141 roleuAlaLeuArgAlaValLeuGlyLeuLeuGlyLeuProAlaAlaAsp 157
   ||| : : : : : |||
452 CGGTGCGCATCCGTCGATCTCCGAACTCTCGCGGTGCCCGAGCCGGAC 501
   ||| : : : : : |||
158 TrpGlyAlaValGlyArgTrpSerArgAspValGlyArgThrLeuAspAr 174
   ||| : : : : : |||
502 CGCGCGCGCTCCCGCTGTGAGC.....GACGC 530
   ||| : : : : : |||
174 gGlyAlaSerAlaGluAspMetArgArgGlyHisAlaAlaIleAlaGluP 191
   ||| : : : : : |||
531 CTTCGTCTCCCGGAGCATCCCGCGCGCGCGCGCATGGCGCGAGA 580
   ||| : : : : : |||
191 heAlaAspTyrValGluArgAlaLeuAlaArgArgArgArgGluGly 207
   ||| : : : : : |||
581 TGAGCGGTATCTCTCCCGGCTCATGCACTCCAGCGCGGGGAGAGAGCG 630
   ||| : : : : : |||
208 GluAspLeuLeuAlaLeuMetLeuAspAlaHisAspArg.....GlyLe 222
   ||| : : : : : |||
631 GAGGACCTGCTCAGCGCGCTGTGCGGAGCGAGCGAGCGCGCTCCG 680
   ||| : : : : : |||
222 umetSerArgAsnGluIleValSerThrValValThrPheIlePheThrG 239
   ||| : : : : : |||
681 GCTGACCTCCGAGAGAGTGTGCGGTATGCGCCACATCCTGCTCTCCG 730
   ||| : : : : : |||
239 LysHisGluThrValAlaSerGlnValGlyAlaValLeuSerLeuLeu 255
   ||| : : : : : |||
731 GGCACGAGACCAAGCTCATCTGATCCGCAACGCGCATGTACGCGCTGCTC 780
   ||| : : : : : |||
256 AlaHisProAspGlnLeuAspLeuLeuArgArgArgProAspLeuAl 272
   ||| : : : : : |||
781 TCGCACCCCGACAGCTGCGCGCGCGCGCGCGCATGACGCTCTGGA 830
   ||| : : : : : |||
272 aglnAlaValalGluCysLeuArgTyrAspProSerValGlnSerSnt 289
   ||| : : : : : |||
831 CGCGCGCGGTGAGGAGATGTTGCGCTACGAGGCGCGGTGGAATCCGGA 880
   ||| : : : : : |||
289 hrArgGlnLeuAspValAsp...ValGlnLeuArgGlyArgArgLeuArg 304
   ||| : : : : : |||
881 CTTACCGCTTCCCGGTGAGCGCGCTGACGCTGACGAGCGAGCGATATCCG 930
   ||| : : : : : |||
305 ArgAspAspValValValLeuAlaGlyAlaAlaAsnArgAspProAr 321
   ||| : : : : : |||
931 GCGGCTTCCGAGCGGTCTGCTGCTGCGCGAGCGCCACCGCACCGCGGA 980
   ||| : : : : : |||
321 GArgTyrTrpAspArgProAspAspPheAspIleGluArgAspProValPro 338
   ||| : : : : : |||
981 GCGCTTCCCGAGCGCGCATCTGACATCCGCGGAGACACCGCGCGCC 1030
   ||| : : : : : |||
338 ermMetSerPheGlyValAlaGlyMetArgTyrCysLeuGlySerTyrLeuAl 354
   ||| : : : : : |||
1031 ATCTGCGCTTGGCGCGAGCATCCACTTCTGCAATCGCGCGCCCTGGCC 1080
   ||| : : : : : |||
355 ArgThrGlnLeuArgAlaAlaVal...AlaAlaLeuAlaArgLeuProG 370

```

```

1081 CGGTGAGAGCCCGGATGCGCTCCGCGCGCTTCGCAACGCTGCCGA 1130
   ||| : : : : : |||
370 yLeuArgLeuGlyCysAlaSerAspAlaLeuAlaTyrGlnProArgThrM 387
   ||| : : : : : |||
1131 CTTGCGCGCTGACGCTCTCCCGCGCACTGCTGTGATCCGAACCGGA 1180
   ||| : : : : : |||
387 ePheArgGlyLeuAlaSerLeuProIleAlaPheThrProGly 401
   ||| : : : : : |||
1181 TGATTCGCGCGCTCAAGCGCTGCGCATCCGCTGCGCGGAGAGA 1224
   ||| : : : : : |||
seq_name: /cgn2_6/prodata/1/ina/6A.COMB.seq-US-09-320-878-21
seq_documentation_block:
: Sequence 21, Application US/09320878A
: Patent No. 6117659
: GENERAL INFORMATION:
: APPLICANT: ASHLEY, Gary
: APPLICANT: BETLACH, Melanie C.
: APPLICANT: BETLACH, Mary C.
: APPLICANT: MCDANIEL, Robert
: APPLICANT: TANG, Li
: TITLE OF INVENTION: RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASE
: FILE REFERENCE: 300622002120
: CURRENT APPLICATION NUMBER: US/09/320,878A
: EARLIER FILING DATE: 1999-05-27
: EARLIER APPLICATION NUMBER: CIP OF 09/141,908
: EARLIER FILING DATE: 1998-08-28
: EARLIER APPLICATION NUMBER: CIP OF 09/073,538
: EARLIER FILING DATE: 1998-05-06
: EARLIER APPLICATION NUMBER: CIP OF 08/846,247
: EARLIER FILING DATE: 1997-04-30
: EARLIER APPLICATION NUMBER: 60/119,139
: EARLIER FILING DATE: 1999-02-08
: EARLIER APPLICATION NUMBER: 60/100,880
: EARLIER FILING DATE: 1998-09-22
: EARLIER APPLICATION NUMBER: 60/087,080
: EARLIER FILING DATE: 1998-05-28
: NUMBER OF SEQ ID NOS: 34
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 21
: LENGTH: 5970
: TYPE: DNA
: ORGANISM: Streptomyces venezuelae
US-09-320-878-21
alignment_scores:
Quality: 538.50 Length: 398
Ratio: 2.216 Gaps: 8
Percent Similarity: 61.055 Percent Identity: 34.673
alignment_block:
US-09-724-797-36 x US-09-320-878-21 ..
Align seg 1/1 to: US-09-320-878-21 from: 1 to: 5970
15 AspValaArgArgAspProTyrProSerTyrHisTrpLeuLeuArgHisAs 31
   ||| : : : : : |||
1422 GATTTCGCGCGCGATTCGATTCGAGCTACGCGAGACGTCGATCCGAGG 1471
   ||| : : : : : |||
31 pProValHisArg.....GlyAlaHisArgValITPTrpValS 44
   ||| : : : : : |||
1472 TCCGCGCCACCGCGGTGCGCACCCCGAGGGGAGCGAGGTGTGCTGTG 1521
   ||| : : : : : |||
44 eArgPheAlaAspValArgAlaValLeuGlyAspGluArgPheAlaArg 60
   ||| : : : : : |||
1522 TCGGCTGACGAGCGCGCGCGCTGCTGCGCATCCCGGTTCAGACAG 1571
   ||| : : : : : |||
61 ThrGlyIleArgArgPheTrpThrAspLeuValGlyProGlyLeuLeuAl 77
   ||| : : : : : |||
1572 GAC.....TGGCGCAACTCCACGACTCCCTGACCGAGGC 1606

```

```
77 aglu...ilevalGlyAspIleIleuPheGlnAspGluProAspHisG 93
||||| :||| :||| :||| :||| :||| :|||
1607 CGAGGGCGGCGTCAACACACATGCTGAGATCCAGCCCGCGGCGACA 1656
93 LysArgLeuArgGlyValValGlyProAlaPheSerProSerAlaLeuArg 109
||||| :||| :||| :||| :||| :||| :|||
1657 CCGCGCTGGCGAGCTGTGGCCCGTGAAGTGCATGCATGCGCGGTGAG 1706
110 ArgLeuGluProValIleAlaGlyThrValAspAspLeuArgProAl 126
||||| :||| :||| :||| :||| :||| :|||
1707 TTGCTGGCGCGCGGCGTCCAGAGATGCTGACGGCTGCTGAGCCCAT 1756
126 AluAlaIAsp...GlyAlaMetAspValValAspGluLeuAlaTyrP 141
||||| :||| :||| :||| :||| :||| :|||
1757 GCTGGCGGCGCGCGAGCGCGCTGATGATGATGATGATGATGATGATG 1806
141 roleuAlaLeuArgAlaValIleuGlyLeuLeuGlyLeuProAlaIAsp 157
||||| :||| :||| :||| :||| :||| :|||
1807 CGCTGGCGATCACCGTGAATCTCCGAATCTCTGCGCTGCGCGAGCGG 1856
158 TrpGlyAlaValGlyArgTrpSerArgAspValGlyArgThrLeuAsp 174
||||| :||| :||| :||| :||| :||| :|||
1857 CGGCGCGCTTCCTCCGCTGAGAC...GAGCG 1885
174 gGlyAlaSerAlaGluAspMetArgArgGlyHisAlaIleAlaGluP 191
||||| :||| :||| :||| :||| :||| :|||
1886 CTTCGCTTCCCGAGCATCCCGCGAGCGCGAGCGCGCATGCGCGAGA 1935
191 heAlaAspTyrValGluArgAlaLeuAlaArgArgArgGlyGly 207
||||| :||| :||| :||| :||| :||| :|||
1936 TGAGCGGCTTCCTCCGCTGATGATGATGATGATGATGATGATGATG 1985
208 GluAspLeuAlaLeuMetLeuAspAlaHisAspArg...GlyLe 222
||||| :||| :||| :||| :||| :||| :|||
1986 GAGGAGCTGTGACGCGCTCGCGGAGCGAGCGAGAGAGAGCGCTCCG 2035
222 uMetSerArgAsnGluIleValSerThrValValThrPheIlePheArg 239
||||| :||| :||| :||| :||| :||| :|||
2036 GCTGACCTCGAGGAGCTGCTCGTATGCGCCACATCTCTGCTGCGG 2085
239 LysHisGluThrValAlaSerGlnValGlyAsnAlaValLeuSerLeu 255
||||| :||| :||| :||| :||| :||| :|||
2086 GAGCAGGAGACACGATCATGATGCGACAGCGCATGACGCTGCTC 2135
256 AlaHisProAspGlnLeuAspLeuLeuArgArgArgProAspLeuAl 272
||||| :||| :||| :||| :||| :||| :|||
2136 TCGACCCCGACAGCTGCGCGCGCGCGCGAGCATGACGCTTGA 2185
272 agAlaValAlaGluIleuSerLeuArgTyrAspProSerValGlnSer 289
||||| :||| :||| :||| :||| :||| :|||
2186 CGGCGCGCTGAGAGATGTGGCTACGAGGCGCGCGGATGCGCGA 2235
289 hrArgGlnLeuAspValAsp...ValGluLeuArgGlyArgArgLeuArg 304
||||| :||| :||| :||| :||| :||| :|||
2236 CTTACCGCTTCCTCCGCTGAGCGCGCTGACGCGAGCTCATCCG 2285
305 ArgAspAspValValValValLeuAlaGlyAlaAlaAsnArgAspPro 321
||||| :||| :||| :||| :||| :||| :|||
2286 GCGGAGTACAGGCTCTGCTGCTGCGCGAGCGCGAGCGCGAGCGG 2335
321 gArgTyrAspArgProAspAspPheAspIleGluArgAspProValPro 338
||||| :||| :||| :||| :||| :||| :|||
2336 GCGGCTTCGCGAGCGCGACCGCTTCGACATCCGCGGAGACCGCGG 2385
338 erMetSerPheGlyAlaGlyMetArgTyrCysLeuGlySerTyrLeuAl 354
||||| :||| :||| :||| :||| :||| :|||
2386 ATTCGCGCTTCGCGAGCGCATCTGCTGATGCGCGCGCGCTTGGC 2435
355 ArgThrGlnLeuArgAlaVal...AlaAlaLeuAlaArgLeuProAl 370
||||| :||| :||| :||| :||| :||| :|||
2436 CGGTTGGAGCGCGGATGCGCGCTTCGCGAGCGCTGCGAGCGG 2485
370 yLeuArgLeuGlyCysAlaSerAspAlaLeuAlaTyrGlnProAlaGln 387
```

```
2486 CCTCGGCTGAGATCTCCCGCGGAGACTGCTGTGATTCGACCGA 2535
387 ePheArgGlyLeuAlaSerLeuProIleAlaPheThrProGly 401
||||| :||| :||| :||| :||| :||| :|||
2536 TGATCCGCGGCTCAAGCGCGCTGATGATGATGATGATGATGATG 2579
seq_name: /cgn2_6/ptodata/1/fna/6B_COMB.seq:us-09-029-603-4
seq_documentation_block:
; Sequence 4, Application US/09029603
; Patent No. 6210935
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Engel, Natalie
; APPLICANT: Bietenhader, Jurg
; APPLICANT: Toupet, Christine
; APPLICANT: Pospishech, Andreas
; TITLE OF INVENTION: Staurosporin Biosynthesis Gene Clusters
; FILE REFERENCE: 4-2055/A/PCT
; CURRENT APPLICATION NUMBER: US/09/029,603
; EARLIER FILING DATE: 1998-03-20
; EARLIER APPLICATION NUMBER: PCT/EP96/03643
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 6085
; TYPE: DNA
; ORGANISM: Streptomyces longisporoflavus
; FEATURE:
; NAME/KEY: misc_RNA
; LOCATION: (378)..(1665)
; OTHER INFORMATION: ORF
; FEATURE:
; NAME/KEY: misc_RNA
; LOCATION: (1747)..(2553)
; OTHER INFORMATION: ORF
; FEATURE:
; NAME/KEY: misc_RNA
; LOCATION: (2593)..(4011)
; OTHER INFORMATION: ORF
; FEATURE:
; NAME/KEY: misc_RNA
; LOCATION: (4013)..(4999)
; OTHER INFORMATION: ORF
; FEATURE:
; NAME/KEY: misc_RNA
; LOCATION: (5071)..(6085)
; OTHER INFORMATION: ORF
US-09-029-603-4
alignment_scores:
Quality: 508.50 Length: 407
Ratio: 2.101 Gaps: 11
Percent Similarity: 59.459 Percent Identity: 36.364
alignment_block:
US-09-724-797-36 x US-09-029-603-4 ..
Align seg 1/1 to: US-09-029-603-4 from: 1 to: 6085
20 ProTyrProSerTyrHisTrpLeuArgHisAspProValHisArg.. 35
||||| :||| :||| :||| :||| :||| :|||
441 CCTTACCGGCTTACCGGCGGAGCGCGCGCGCTCATCGGAC 490
36 .....GlyAlaHisArgValTrpTyrValSerArgPheAla 48
||||| :||| :||| :||| :||| :||| :|||
491 GCGGCGGCGCGCGGAGAACCTGACACCTACTACTGTTCACTACGACG 540
48 sPValArgAlaValIleuGlyAspGluArgPheAlaArgThrGlyLeuArg 64
||||| :||| :||| :||| :||| :||| :|||
```



```
46716 .....ACCGATCCCGCCAGCCCGGATGTTCCCAACCCGCGCGGAGC 46758
      ||||| : : : ||| : : : : : : :
81 LyspIleIleLeuPheGlnAspGluProAspHisGlyArgLeuArgGly 97
      ||||| : : : ||||| ||||| ||||| ||||| |||||
46759 CGGACGGCGTCGCGCCAGAGCCCGCGGACCAACCCGCGTGGGCGG 46808
      ||||| : : : ||||| : : : : : : :
98 ValValGlyProAlaPheSerProSerAlaLeuArgArgLeuGluProVa 114
      : : : ||||| : : : : : : :
46809 CTGGTGGGCAAGGCTTTCAGCGAGCGCGGTGGAGAGATGGCGGCCCG 46858
      : : : ||||| : : : : : : :
114 LlleAlaGlyThrValAspAspLeuArgProAlaLeuAlaArgGlyA 131
      : : : ||||| : : : ||||| : : : |||||
46859 TGTCCGCTCCCTCGTCACTCCCTGCTGAGACATGGTGGCGCAGGTT 46908
      : : : ||||| : : : ||||| : : : |||||
131 La...MetAspValValAspGluLeuAlaTyrProLeuAlaLeuArgLa 146
      : : : ||||| : : : ||||| ||||| : : : : :
46909 CACCGCGGACCTGTGCGAGTTCGCGCTTCCCTCCCGCGCGGCTC 46958
      : : : ||||| : : : ||||| : : : |||||
147 ValLeuGlyLeuLeuGlyLeuProAlaAlaAspTyrGlyAlaValGlyAr 163
      : : : ||||| : : : ||||| : : : |||||
46959 ATCGCGCACTGCTGCGCGCTGCGCTTGGAGGAC..... 46991
      : : : ||||| : : : ||||| : : : |||||
163 gTTPSerArgAspValGlyArgThrLeuAsp..... 173
      ||||| : : : ||||| : : : ||||| : : : |||||
46992 .....CGGACCTGTCCGAGCTTTCACGCGCATGCTCTCTCGA 47034
      ||||| : : : ||||| : : : ||||| : : : |||||
174 ..ArgGlyAlaSerAlaGluAspMetArgGlyHisAlaAlaLea 189
      ||| : : : ||||| : : : ||||| : : : |||||
47035 CCGCGCTCACCGCGCGGAGATACAGCG.....GTCCAGCAG 47072
      : : : ||||| : : : ||||| : : : |||||
190 GluPheAlaAspTyrValGluArgAlaLeuAlaArgArgArgGluG 206
      : : : ||||| : : : ||||| : : : |||||
47073 GACTTCATGGTCTACATGAGCGCTGCTGCCAGCGCGCGAGCGCCC 47122
      : : : ||||| : : : ||||| : : : |||||
206 yGlyGluAspLeuLeu..AlaLeuMetLeuAspAlaHisAspArgGlyL 222
      ||||| : : : ||||| : : : ||||| : : : |||||
47123 CACCGAGACCTGCTCGCGCGCTCCGCTCCGCAACGCAACGAGCAGC 47172
      : : : ||||| : : : ||||| : : : |||||
222 eumetSerArgAsnGluIleValSerThrValValThrPheLeuPheThr 238
      : : : ||||| : : : ||||| : : : ||||| : : : |||||
47173 ACCGACCAAGGCGGAGATGTCACATGCGGGGTGAGTGTCTCATGCGG 47222
      : : : ||||| : : : ||||| : : : |||||
239 GlyHisGluThrValAlaSerGlnValGlyAsnAlaValLeuSerLeu 255
      ||||| : : : ||||| : : : ||||| : : : |||||
47223 GCGCAGAGACGTCGCGTCAACGATCACCACTGTCACCTGCTGCT 47272
      : : : ||||| : : : ||||| : : : |||||
255 uAlaHisProAspGlnLeuAspLeuLeuArgArgArgProAspLeuLeuA 272
      : : : ||||| : : : ||||| : : : ||||| : : : |||||
47273 GACCGAGCGCAAGCGCTACAGTCTGCTGCCGCGCACCGCGCTCGTGC 47322
      : : : ||||| : : : ||||| : : : |||||
272 laGlnAlaValGluGluCysLeuArgTyrAspProSerValGlnSerAsn 288
      ||||| : : : ||||| : : : ||||| : : : |||||
47323 CCGCGCGGTGGAGAGATGTCGCGGTACACACCGCTGTGTCCGCGCGC 47372
      : : : ||||| : : : ||||| : : : |||||
289 Thr.....ArgGlnLeuAspValAspValGluLeuArgGlyArgGlyLe 303
      : : : ||||| : : : ||||| : : : ||||| : : : |||||
47373 AGCTTCCTCGCTGCGCGCACCGAGCTGCGAGCTGACACCGCTGACCGT 47422
      : : : ||||| : : : ||||| : : : |||||
303 uArgArgAspAspValValValValLeuAlaGluValAlaAsnArgAsp 320
      : : : ||||| : : : ||||| : : : ||||| : : : |||||
47423 GCGGCGCGGGAGACCTGCGTGTCTCATTCTGCGTGGCGCAACGGGAGC 47472
      : : : ||||| : : : ||||| : : : ||||| : : : |||||
320 roArgArgTyrAspArgProAspAspPheAspIleGluArgAspProVal 336
      : : : ||||| : : : ||||| : : : ||||| : : : |||||
47473 AGGAGGCTTTCGACACGCGGAGACCTGACTTCCACCGTGAAGCGCAAC 47522
      : : : ||||| : : : ||||| : : : ||||| : : : |||||
337 ProSerMetSerPheGlyAlaGlyMetArgTyrCysLeuGlySerTyrLe 353
      ||| : : : ||||| : : : ||||| : : : ||||| : : : |||||
47523 CCGGACATACGCTTGGCGCAGGAGCGACCACTGATCGGCGCGCAACT 47572
      : : : ||||| : : : ||||| : : : ||||| : : : |||||
353 uAlaArgThrGlnLeuArgAlaAlaValAlaLeuAla...ArgLeuP 369
      : : : ||||| : : : ||||| : : : ||||| : : : |||||
```

```
47573 GCGCGACTGGAACCTCAGAGAGCCCTGTCCGCCCTGTCCGCGGCTTCC 47622
      ||||| : : : ||||| : : : ||||| : : : |||||
369 roGlyLeuArgLeuGlyCysAlaSerAspAlaLeuAlaTyrGlnProArg 385
      ||||| : : : ||||| : : : ||||| : : : |||||
47623 CCACCTGATGTGGCGAGCGGTCGCGGAGCTGAAGTGAAGCAGGCG 47672
      : : : ||||| : : : ||||| : : : ||||| : : : |||||
386 ThrMetPheArgGlyLeuAlaSer.....LeuProI 396
      : : : ||||| : : : ||||| : : : ||||| : : : |||||
47673 ATGCTGATCCCGGAGATGGAAGCGCATGTCTCTGTTGACGCGCGGC 47722
      : : : ||||| : : : ||||| : : : ||||| : : : |||||
396 eAlaPheThrProGlyGly 402
      ||| : : : ||||| : : : ||||| : : : |||||
47723 CGCGCGCGCGCGCGCGG 47741
      : : : ||||| : : : ||||| : : : ||||| : : : |||||

seq_name: /cgn2_6/prodata/1/lna/Backfile1.seq:5212296-17
seq_documentation_block:
: Patent No. 5212296
: APPLICANT: DEAN, CAROLINE; HARDER, PATRICIA A.; LETO, KENNETH
: J.; O'KEEFE, DANIEL P.; OMER, CHARLES A.; ROMESSEK, JAMES A.
: TEPPERMAN, JAMES M.
: TITLE OF INVENTION: EXPRESSION OF HERBICIDE METABOLIZING
: CYTOCHROMES
: NUMBER OF SEQUENCES: 19
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/07/569,781
: FILING DATE: 23-AUG-1990
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 464,499
: FILING DATE: 12-JAN-1990
: APPLICATION NUMBER: 405,605
: FILING DATE: 11-SEP-1989
: SEQ ID NO: 17
: LENGTH: 1212
5212296-17

alignment_scores:
      Quality: 454.00      Length: 389
      Ratio: 1.924      Gaps: 10
      Percent Similarity: 60.668      Percent Identity: 34.447

alignment_block:
US-09-724-797-36 x 5212296-17 ..
Align seg 1/1 to: 5212296-17 from: 1 to: 1212

19 AspProTyrProSerTyrHisThrPheLeuArgHisAspProValHisAr 35
      ||||| : : : ||||| : : : ||||| : : : |||||
76 GACCGCGCGCGCGAGCTCACCGAGCGCGCGCGCACCGAGCGGTCACCGG 125
      : : : ||||| : : : ||||| : : : ||||| : : : |||||
35 gGly.....AlaHisArgValTyrTyrValSerArgPheAla 48
      : : : ||||| : : : ||||| : : : ||||| : : : |||||
126 GCGCACCTCTGGGAGCGGCTCTCTCTGCTGTGTGAGCGCGCATCAG 175
      : : : ||||| : : : ||||| : : : ||||| : : : |||||
48 spValArgAlaValLeuGlyAspGluArgPheAlaArgThrGlyLeArg 64
      ||||| : : : ||||| : : : ||||| : : : ||||| : : : |||||
176 AGCTCCGCGCGGCTCTCTCTGCGGACCGCGCTTTCAGCGCGCACCGCCACCGC 225
      : : : ||||| : : : ||||| : : : ||||| : : : |||||
65 ArgPheThrPheAspLeuValGlyTyrGlyLeuLeuAlaGluIleValG 81
      : : : ||||| : : : ||||| : : : ||||| : : : |||||
226 ACCGCTTCCCTTCTGACCGCGCGGCG.....CGGAGATCATCGG 269
      : : : ||||| : : : ||||| : : : ||||| : : : |||||
81 Y.....AspIleIleLeuPheGlnAspGluProAspHisGlyArgLeu 96
      : : : ||||| : : : ||||| : : : ||||| : : : |||||
270 CACCAACCGGACCTTCTGCGCATGAGCAGCGCGGAGCAGCGCGCATGCG 319
      : : : ||||| : : : ||||| : : : ||||| : : : |||||
96 rGlyValValGlyProAlaPheSerProSerAlaLeuArgArgLeuGlu 112
      ||| : : : ||||| : : : ||||| : : : ||||| : : : |||||
320 GCGGATGCTCACCGCGGACTTATGCTCAAGAAGTTCAGAGCGATGCGC 369
      : : : ||||| : : : ||||| : : : ||||| : : : |||||
113 ProValIleAlaGlyThrValAspAspLeuLeu...ArgProAlaLeuAl 128
      ||| : : : ||||| : : : ||||| : : : ||||| : : : |||||
```

```
370 CCCAGGTGCAGCGCCCTCCGACGACCTGTCGACCGGATGACCACCGG 419
128 aargllylmetaspvalvalaspglutlalatyrproleuallaleua 145
420 ACGCACTCCCGCCGCTGGTACCGAGTTGGCGGTGGCGGCTGCC 469
145 rglavallaleu glyleu glyleu prolaalasprrp glyalaval 161
470 TGGTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 519
162 GYATGTPPserArgAspValGlyArgThrleuasparglyAlaSerAl 178
520 CAGGAGCCGACGGGGTCTG.....CTCACCTCGCGGTCCACCTCC 560
178 agluspmetargarglyHisAlaAlaIlealagluphealasprry 195
561 CGAGGAGATCGCG.....GCCGCCAGGACGAGTTGGTGGAGTACC 601
195 aIGluArgAlaLeuAlaArgArgArgGlyGlyGlyGlyGlyGlyGly 211
602 TCGCCCGG...CTCGCCCGACCAAGCGGAGCGCGCGCGCGCGCATC 648
212 AlaleuMetleuaspAlaHisAspArgGlyLeuMetSerArgAsnGlu 228
649 ATCAGCGCGCTGCTGCC.....CGCGCGAGCTCGACGACCAACCAT 692
228 eValSerThrValValThrPheIlePheThrGlyHisGlyThrValAla 245
693 CGCCACCAATGCGAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 742
245 ercInValGlyAsnAlaValLeuSerLeuAlaHisProaspGlnLeu 261
743 ACATGACCGCGCTCTCCACCTGCTGCTGCTGCTGCTGCTGCTGCTG 792
262 AspleuLeuArgArgArgProaspLeuAlaGlnAlaValGluGly 278
793 GCCCGGTGCGCGCGACCGCGCTGCTGCTGCTGCTGCTGCTGCTGCT 842
278 slauatgTyrAspProserValGlnSerAsnThrArgGlnLeuAspVal 294
843 GCTGCGTACCTGACATCGTCACAAAGCGCTTCCCGCGATGCCACCG 892
295 .AspValGluLeuArgGlyArgArgLeuArgArgAspAspValVal 310
893 AGGACGCTCATCGCGCGCGACCATCGCGCGCGCGCGCGCTGCTG 942
311 ValLeuAlaGlyAlaAlaAsnArgAspProArgArgTyrAspArgPro 327
943 TGCATGATCAAGCTCCGCAACCGGAGCGCGAGTGTCCCGCGCGCA 992
327 PAspPheAspIleGluArgAspProValProSerMetSerPheGlyAla 344
993 CGACCTGAGTGGCGCGCGCGCGCGCGCGCGCGCGCGCTTGGCTGCG 1042
344 lYmetArgTyrCysLeuGlySerTyrLeuAlaArgThrGlnLeuAla 360
1043 GCGTCACAGTGGCTGGGACAGCGCTTGGCGCGGAGTGCACATC 1092
361 AlAlaVal...AlaAlaLeuAlaArgLeuProGlyLeuArgLeuGly 376
1093 GCCATCAAAAGCTGCGCGCGCTGCGCGCGCGCGCGCGCTGCGCGG 1142
376 aSerAspAlaLeuAlaTyrGlnProArgThrMetPheArgGlyLeuAla 393
1143 CCAGAGAGAGATCCCGTTCCGCGCGCGCGCGCATGCGCATCGGGGT 1192
393 erleuProIleAlaPhe 398
1193 CGTGGCGATGCGCTGG 1209
seq_name: /cgn2_6/ptodata/1/ina/5B_COMB.seq:us-09-103-840A-2
seq_documentation_block:
```

```
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; FILE REFERENCE: 24366-2007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4403765
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: CDC 1551
; OTHER INFORMATION: "n" bases at various positions throughout the sequence
; OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2
```

alignment_scores: Quality: 446.00 Length: 414

Percent Similarity: 57.005 Gaps: 11

Percent Identity: 31.401

Alignment_block: US-09-724-797-36 x US-09-103-840A-2/rev ..

```
Align seg 1/1 to reverse of: US-09-103-840A-2 from: 1 to: 4403765
6 valthrAlaPheaspProthrAspAlaAspValArgArgAspProTyr 22
2541149CTCACCGAGCTTGGACCGCTTGGATCCCGCGATGCGCGTGTACCTTACC 2541100
22 oSerThrIstPleuLeuArgHisAspProValHisArgGlyAlaHis 39
2541099GATATATGAGAGCTGTGGCGCGCGCGCGCGGTGACATCAACCCCAAGC 2541050
39 rg...ValThrPyrValSerArgPheAlaAspValArg.....AlaVal 52
2541049GCGAGCTCATATTGAGCCGATACGCCGACAGTCCGCGAGCGCTGCGGT 2541000
53 LeuGlyAspGluArgPheAlaArgThrGlyLe.....ArgArgPheTr 67
2540999AACCAACGACGCTGTCCAGCGCGCGCGCGCGCTTCTCAGCGGGGTG 2540950
67 pThrAspLeuValGlyProGlyLeuLeuAlaGluIleValGlyAspIle 84
2540949GCTGCGCTTTCGCGAGC..... 2540931
84 leuLeuPheGlnAspGluProAspHisGlyArgLeuArgGlyValValGly 100
2540930.....TCCATCCGCGCGCGCGCGCGCGCGCGCGCGCGCGTGGCA 2540889
101 ProAlaPheSerProSerAlaLeuArgArgLeuGluProValIleAlaG 117
2540888CCAGATATGCGCGCGCGCTTGGAGAGTGGCGCGCGCGCGCGCGTGG 2540839
117 yThrValAspAspLeuArgProAlaLeuAlaArgGlyAlaMetAsp 134
2540838GCTTCCCGAGAACTGTGCGCGGTACTGTACCAAGACCGCGCGGACG 2540789
134 aValAspGluLeuAlaTyrProleuAlaLeuArgAlaValLeuGly 150
2540788TGTCTCCACCGTGGCGCGCGCGCGCGCGCGCGCGCGCGGTGG 2540739
151 LeuGlyLeuProAlaAlaAspTrpGlyAlaValGlyArgTrpSerArg 167
```


[illegible]

Alignment	Quality	427.00	Length	393
Percent Similarity	Ratio	1.825	Gaps	12
		59.542	Percent Identity	34.097
Alignment block:				
US-09-724-736 x 5212296-8 ...				
Align seg 1/1 to: 5212296-8 from: 1 to: 1998				
19	ASPRQTYRPROSERTYHISITRPLEUQAIGHISAPROVALHISAR	35		
270	GACCCCGGCGCCGACGTACCGAGCGGCGCGCACCGACGGGTACCG	319		
35	gely.....AIAHSARGVALITRPTYVALISERARGPHEALAA	48		
320	GGCCACCCSTTGGGAGCGGTCTCTGCTGGCTGTATACCGCATCAGG	369		
48	SPVALARGALVALLEUCLIVASPGIUAARGPHEALAAITGHTGLILEARG	64		
370	ACGTCCGGCGGTCTCGGCGACCGCGGCTTACACGGCGACGCCACCG	419		
65	ARGPHEITRPHASPLEUVALGLYPROGLILEUQIACTUILLEVALGI	81		
420	ACCGGCTTCCCTTCTGACCGCGCGCGC.....CGCGAGATCATCGG	463		
81	y.....ASPILEILEUPHEGINSAPGILPROASPHISGLIARGHEUA	96		
464	CACCAACCCGACCTTCCTGGCATGGAGACACCGCGACCGCGCATGCG	513		
96	RGILVALVALGLYPROALAPHESETPROSERALALAEUARGARGLEU	112		
514	CGCGATGCTACCGCGCATTCATTCGTCAAGAAAGTCAAGGCGATGCG	563		
113	PROVALILEAGLYTHVALASPSPLEU...ARGPROALAEUAL	128		
564	CCCGAGGTGCACCGCTCGCGACGACCTGTCCACCGATGACACACCG	613		
128	ARGGLYAMETASPRVALIVASPGIULEALITUPROBLEUALAEUA	145		
614	ACGCACTCCGCGCGATGGTGTACCGAATTCGGCTCCCTGCGCTCC	663		
145	rgAlaValLeuclYleuclYleuPProAlaAlaAspTgIyAlaVal	161		
664	TGTTATTCGCTGCTGCTGGCGGCTCCCTGACGAGCAACGGTCTTC	713		
162	GLYATGTTPSERATGASPVALGLYARGTHILEUASPRGLYALASERAI	178		
714	CAGGACGCGACCGGCTGCTG.....CTCACCTCGGCTCCACTCC	754		
178	agIuAspMetAdgValgIyHISAlaAlaIlealGluPhAlaAspTgY	195		
755	CGAGGAAGTCCG.....GGCGCCAGACGACCATTCGTGGAGTACC	795		
195	alGluAlgAlaLeuAlaIaTgArgTgArgTgIyGluGluAspLeucl	211		
796	TGGCCCGG...CTCGCCGACCAAGACGGAGCGCGCGAGACGGCATC	842		
212	AlaLeuMetLeuAspAlaHISAPRGILYLEUMETSERARGANSGLUI	228		
843	ATCAGACGCGCTGTGCTG.....CGCGCGAGCTGCAGACACCCAGAT	866		
228	eYalSerTgIValIVAlIThrPheIlePheThcIyHIS...GluThYValA	244		
887	CCCCACCATGGGACCGCTGTTGCTGTGCTGGCGCGCGGACCTCGACG	936		
244	leSerGluValGluSnaIValValLeuSerLeuAlaHISPROASPGIn	260		
937	ACACCCAGATGCGCCACATGGGACTG...CTGCTGGCCACACCCGACAA	983		
261	LeuAspLeuAlaArgTgArgTgProAspLeuclAlaGluValGluG	277		

Query	Subject	Quality	Ratio	Percent Similarity	Length	Gaps	Percent Identity
19	AspProTyTProSerTyHisTrpLeuArgHisAspProValHisAr	427.00	1.825	59.542	393	12	34.097
270	GACCCCGCCGCCAGCTCACCGAGCGCGCCGACCAACCGGACCCG	319					
35	ggly.....AlaHisArgValTrpTyValSerArgPheAla	48					
320	GCCACCCCTCTGGAGCGCTCTCTGCTGGCTGTGTACGCGCATACG	369					
48	spValArgAlaValLeuGlyAspGluArgPheAlaArgThrGlyIle	64					
370	ACGTCCGCGCGGCTCCGCGACCGCGCTTCACGCGCCGACCCACG	419					
65	ArgPheTrpThrAspLeuValGlyProGlyLeuLeuAlaGluIle	81					
420	ACCGGCTCCCTCTCTACCGCGCGGCG.....CGCGAGATCATGG	463					
81	Y.....AspIleIleLeuPheGlnAspGluProAspHisGlyArg	96					
464	CACCAACCGACACTTCTCTGCGATGGACGACCGGACGACGCGCATG	513					
96	rgGlyValValGlyProAlaPheSerProSerAlaLeuArgGlyLeu	112					
514	GCGGATGTCTACCGCGCATCATGTCAGAGAGTCCGAGCGCATCGC	563					
113	ProValIleAlaGlyThrValAspLeuLeu...ArgProAlaLeuAl	128					
564	CCGAGAGTGCAGCGCTCTGCGACGACCTGTGTCCAGCGATAGCACCG	613					
128	aarGlyAlaMetAspValValAspGluLeuAlaTyTProLeuAlaLeu	145					
614	ACGCACCTCCGCGACCTGTGTACAGAGTGTGCGCTCGCTCGCTGCC	663					
145	rgAlaValLeuGlyLeuGlyLeuGlyProAlaAlaAspTrpGlyAla	161					
664	TGTGTATGTCTCTGCTGCTGCGCGCTCTCTACGAGAGCACGCGTTCT	713					
162	GlyArgTrpSerArgAspValGlyArgThrLeuAspArgGlyAlaSer	178					
714	CAGAGCGGACGCGGCTGTG.....CTCACCTCGCGGTCTCACTCC	754					
178	agLysMetLeuArgGlyGlyHisAlaAlaIleGluPheAlaAspTyT	195					
755	CGAGGAAGTCCG.....GCCGCCAGACAGAGTGTGTGGAGTACC	795					
195	alGlyArgAlaLeuAlaArgGlyArgGlyGlyGlyGlyGlyAspLeu	211					
796	TGCGCCGG...CTGCGCGGACCAAGCGGAGCGCGGACGACGCTATC	842					
212	AlaLeuMetLeuAspAlaHisAspArgGlyLeuMetSerArgAsnGlu	128					
843	ATCAGCGCGCTGTGCTCC.....CGCGGACGCTCGACGACACCGAT	886					
228	evalSerTrpValValThrPheIlePheThrGlyHis...GluThrVal	244					
887	CGGCACCATGGGACGCTTCTGTGTGTCGCGCGCGCGGACGCTGCAG	936					
244	IsaSerGlnValGlyAsnAlaValLeuSerLeuLeuAlaHisProAspGln	260					
937	ACACCCAGATGCGACCATGGAGTGC...CTGTGTGCGCACCGCGACCA	963					
661	LeuAspLeuLeuValArgArgTrpArgAspLeuLeuAlaGlnAlaValGlu	277					

```
984 CTCGCCGCTGGCGCCGGAACCCGGCTGTCAAGGCGCCGTCGAGA 1033
277 ucysleuargtyraspproservalginserasnthrarglnleuaspy 294
1034 GCTGCTGCTACCTGACATCGTGCACACAGGCGCTTCCCGGATCCCA 1083
294 a...aspvalgluleuarglyargargleuargraspaspyalval 309
1084 CCGAGACGTGCTCATCGGCGCGCACCATCGCCGCGGAGGCGCTC 1133
310 valvalleualaglyalaalaasnargaspproargargtyrasparpr 326
1134 CTGGGCATGATCAGCTCCGCCAACGCGGACGCGAGGTGTCCCGCGCG 1183
326 oasppasphaspillegluargaspprovalprosermetserpneglya 343
1184 CGACGACCTCGACGTGGCGCGGACGCGCGCCACGCTGGCCTTGCGGT 1233
343 laglymetargtyrcysleuglysettyrleualaargthrlnleuatg 359
1234 TCGGCGTCCACACAGTGGCTGGGACGCGCTGGCCAGGGGTGAGCTCCAG 1283
360 AlaAlaVal...AlaAlaLeualaArgLeuProGlyLeuArgLeuGlyCy 375
1284 ATGCCCATCGAAGCGTGTGCGCGCGCTGGCGGACCTGGCGTGGCGGT 1333
375 salaseraspalaleualaargthrlnprcargthrmetpheatrgglyleua 392
1334 GCCCCAGAGAGATCCGCTTCCGCGGACATGGCATCTACGGGGTCC 1383
392 laserleuproilealaphethrprogly 401
1384 ACTGCTGCCGATCGCTGTGATGCCGCGG 1412
seq_name: /cgn2_6/ptodata/1/ina/backfile1.seq:5212296-16
seq_documentation_block:
; Patent No. 5212296
; APPLICANT: DEAN, CAROLINE; HARDER, PATRICIA A.; LERO, KENNETH
; J.; O'KEEFE, DANIEL P.; OMER, CHARLES A.; ROMESSER, JAMES A.
; TERPPEMAN, JAMES M.
; TITLE OF INVENTION: EXPRESSION OF HERBICIDE METABOLIZING
; CYTOCHROMES
; NUMBER OF SEQUENCES: 19
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/569,781
; FILING DATE: 23-AUG-1990
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: 464,499
; FILING DATE: 12-JAN-1990
; APPLICATION NUMBER: 405,605
; FILING DATE: 11-SEP-1989
; SEQ ID NO:16
; LENGTH: 1221
5212296-16
alignment_scores:
Quality: 417.50 Length: 421
Ratio: 1.625 Gaps: 13
Percent Similarity: 61.045 Percent Identity: 30.166
alignment_block:
US-09-724-797-36 x 5212296-16 ..
Align seg 1/1 to: 5212296-16 from: 1 to: 1221
2 leuValaspAlaValThrAlaPheaspProThrAspAlaAspyal.....16
1 ATGACCGATACGCCGACGACGCCGACACGACGACGACGACCGCTTCC 50
17 .....ArgArgaspProTyr.....ProSerTyrHisTrpLeuLeua 29
||||| ||||||| ||| |||
```

```
51 GAGCAACCGGAGCTGTCCTACCATCGGACGCGCTACCGCCAGCTCC 100
29 rghis.....AspProValHisArg.....GlyAlaHisArg 39
101 GGGACACCCCGGCGCCCTGACCCGGGTGACGCTTACGACGCGCTGAG 150
40 ValTyrPtyrValSerArgPheAlaAspValArgAlaValLeuGlyAsp 56
151 GCGTGGGTGTGACCAAGACGAGGCGCGCGCAACTGCTGGGAGCCC 200
56 uArgPheAla.....ArgThrGlyLeuArg.....A 65
201 CCGGCTGCTCTCCACCGGACGAGCAACTTCCCGCCAGTCACCGC 250
65 rghPheTrpThrAspLeuValGlyProGlyLeuLeuAlaGluLeuValGly 81
251 GCTTCGAGGCGCTCCGAGAGCGCG..... 276
82 AspIleIleLeuPheGlnAspGluProAspHisGlyArgLeuArgGly 98
277 CAGCGCTTCATCGCCTGACCGCCGCGGACGCGCACCGGACCGCGCGGAT 326
98 lValGlyProAlaPheSerProSerAlaLeuArgArgLeuGluProVal 115
327 GAGCATCAGGAGATTCACTGATGAGGATCAAGGCGATGCGCCCGAG 376
115 leaAlaGlyThrValAspAspLeuLeuArgProAlaLeuAlaArgGly... 130
377 TCGAGGAGGTGTGCACGCGCTCTCGACAGAGATGCTGCGCCCGCGCG 426
131 AlaMetAspValAlaAspGluLeuAlaTyrProLeuAlaLeuArgAla 147
427 ACCGCGCACTGGTCACTGATCGCGCTGCGCGTCCCTCATGATGTAT 476
147 lLeuGlyLeuLeuGlyLeuProAlaAlaAspTrpGlyAlaValGlyArgT 164
477 TGCCGCACTCTCGCGCTGCGCTTACCGCGAC.....CACGAGT 514
164 rPserArgAspValGlyArgThrLeuAspArgGlyAlaSerAlaGluAsp 180
515 TCTTCCAGACGCGGACGACGCGCTG.....GTGAGTCCACGAGC 555
181 MetArgArgGlyHisAlaAlaIleAlaGluPheAlaAspTyrValGluArg 197
556 GCGCAGAGCGCGCTCACCGCGGAGACGCTGCGCTTACCTGAGACG 605
197 gAlaLeuAlaArgArgArgGlyGlyGlyAlaAspLeu.....AlaAla 213
606 CTTATCACCCAGTTCAGACGACGCGGCGCGGCTGAGGCGCTC 655
213 eukleuAspAlaHisAspArgGlyLeuMetSerArgAsnGluIleVal 229
656 TGCTGCGCGACGCTGCGCCACGCGAGATCACCGGTGAGAGACGATC 705
230 SerThrValAlaThrPheIlePheThrGlyHisGluThrValAlaSerG 246
706 TCCACGCGATGCTGCTCTCATCGCGCGGACGAGACCGGCTCGAT 755
246 nValGlyAsnAlaValLeuSerLeuLeuAlaHisProAspGlnLeuAsp 263
756 GACCTCCCTCAGCGTATCACCTGTGACACCGCGGAGATGACGCGG 805
263 euLeuArgArgArgProAspLeuLeuAlaGlnAlaValGlyGlySer 279
806 CCTGCGCGCGGACCGGCTGCTGCGCGCGCGCGGTGAGAGAACTGCTC 855
280 ArgTyrAspProSerValGln...SerAsnThrArgGlnLeuAspValAs 295
856 CGTACTCTCGCATCGCGCATCGCGGCGCGCGGTGCGCACCGCGGGA 905
295 pValGluLeuArgGlyArgArgLeuArgArgAspAspValValValVal 312
906 CATGAGGTGAGGCGCATCTACGCGCGCGCGCGGAGGCGGTGATGCTG 955
```



```
345  LArgTyrCysLeuGlySerTyrLeuAlaArgThrGlnLeuAlaAlaIay 362
      :::: |||||||::: ||||||| :::::
1183 GCACGAGTGGCTGGCCAGACCTGCCCGGCTGGAGCTGGAGGTCAATCC 1232
      :: ||||||| :::::
362  AlAlaAlaLeuAla...ArgLeuProGlyLeuArgLeuGlyCysAlaSer 377
      :: ||||||| :::::
1233 TCAACGCCCTCATGACCGGCTCCGACGCTGCACTGCCCTCCCGCTC 1282
      :: ||||||| :::::
378  AsPAlaLeuAlaTyrGlnProArgThrMetPheArgGlyLeuAlaSerIe 394
      :: ||||||| :::::
1283 GAGAGTTGGTGGCTGGCGCGGTACGACGATCAGGCGCTCAACGACT 1332
      :: ||||||| :::::
394  uProIleAlaPhe 398
      :: ||||||| :::::
1333 CCGGTCACCTGG 1345

seq_name: /cgn2_6/ptodata/1/ina/5A_COMB.seq:us-08-102-863-10

seq_documentation_block:
; Sequence 10, Application US/08102863
; Patent No. 5466590
; GENERAL INFORMATION:
; APPLICANT: SARIASIANI, SIMA
; TITLE OF INVENTION: CONSTITUTIVE
; TITLE OF INVENTION: EXPRESSION OF P450SOY
; TITLE OF INVENTION: AND FERREDOXIN-SOY IN
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: E. I. DU PONT DE NEMOURS
; STREET: 1007 MARKET STREET
; CITY: WILMINGTON
; STATE: DELAWARE
; COUNTRY: USA
; ZIP: 19898
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0,
; SOFTWARE: Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/102,863
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/807,001
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: GALLEGOS, R. THOMAS
; REGISTRATION NUMBER: 32,692
; REFERENCE/DOCKET NUMBER: CR-9000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 302-892-7342
; TELEFAX: 302-892-7949
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1735 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-102-863-10

alignment_scores:
Quality: 410.00 Length: 428
Ratio: 1.760 Gaps: 10
Percent Similarity: 54.439 Percent Identity: 29.439

alignment_block:
US-09-724-797-36 x US-08-102-863-10
```

```
Align seq 1/1 to: US-08-102-863-10 from: 1 to: 1735

10  AsPProThrAspAlaAspValAlaArgAspProTyrProSer..... 23
      |||||||
223  GACCCACAC.....TCCCGGCCCCCGGACAGCTTCCTT 254
      :::::
24  .....TyrHisTrpLeuAlaArgHisAspProv 33
255  CCGGACGAGACCGGGGTGCCCTTACCAACCCGCCCGGGGTACGACGCC 304
      :::::
33  AlHisArgGly.....AlaHisArg 39
      :::::
305  TCGCGGAGGCGCCCGCTGAGCGGGTCACCTTTCGACGAGACGCCG 354
      :::::
40  ValTrpTyrValSerArgPheAlaAspValAlaValAlaValLeuGlyAsp.. 55
      |||||||
355  GTCTGGGCGGTACCGGGACGCGCTGCGCTGCGCTACTGCGGAGACC 404
      :::::
56  .....GluA 57
405  GCGGCTCTCCACCGACCGACCGACCGGACTTCCCGTCCGCGCGGAC 454
      :::::
57  rPheAlaArgThrGlyIleArgArgPheTrpThrAspLeuValGlyPro 73
      |||||||
455  GGTTCGCGGCGCGACGCGCGCGCGCTC..... 483
74  GlyLeuAlaGluIleValGlyAspIleIleLeuPheGlnAspGluPr 90
      :::::
484  .....GCTCTGCTGGGCTGACGACCC 506
90  oAsPHisGlyArgLeuArgGlyValAlaGlyProAlaPheSerProSerA 107
      :::::
507  CGACACACACACACCGACGCGAGATGCTCATCCGACCTTCTGCTGAAC 556
107  LAlaLeuArgArgLeuGluProValIleAlaGlyThrValAspAspLeu 123
      :::::
557  GGATCGGCGGCTCCGCCGCTATCCAGAGACCGTGGACGGCTCTTC 606
124  ArgProAlaLeuAlaArgGly...AlaMetAspValAlaAspGluLeuAl 139
      :::::
607  GACCGATGAGCGGACGACGAGCGCCCGGACGTAAGCGCGCTTCGC 656
139  aTyrProLeuAlaLeuAlaGluValLeuGlyLeuGlyLeuProAlaA 156
      :::::
657  CCTGCGGCGCGCTCGATGCTGATGCTGCTGCTGCGCGGCTTACG 706
156  LAsPTrpGlyAlaValAlaGlyArgTrpSerArgAspValGlyArgThrLeu 172
      :::::
707  CCGACACACGG.....TTCCTGAGGACGCTCGGACGACTC 744
173  AsPArgGlyAlaSerAlaGluAspMetArgArgGlyHisAlaAlaIleAl 189
      :::::
745  CTGGCGGCGCGGAGCGGACGATGTGAACG.....GCCGCGA 785
189  aGluPheAlaAspTyrValGluArgAlaLeuAlaArgArgArgGluG 206
      :::::
786  CGAAGCTGAGAGAGTACTGGGCGCGCTGATCGACCCGACAGGGGAGAC 835
836  CGGGTGAAGCGGCTCTCGACGAGCTGATCCACCGGAGACCGCGGAGGA 885
222  LeuMetSerArgAsnGluIleValSerThrValValThrPheIlePheh 238
      :::::
886  CCGGTGACGCGCGAACAGCTGGTCCCTTCCGCTATCCTGCTCATGCC 935
238  rGlyHisGluThrValAlaSerGlnValGlyAsnAlaValIleSerLeu 255
936  CCGGACAGAGACGACGCGGACATGATCTGCTGCGGACGTTACGCTGC 985
255  eAlaHisProAspGlnLeuAspLeuAlaArgArgProAspLeu 271
      :::::
```

```

986 TGAGCACCCCGAACAGCTGGCGGCGCTCGGGCCGCGGGAGCAGCACCC 1035
272 AAGLAlaValaIglucIucysIleuArgTyRasPProSerValaIgluSerAs 288
1036 GCCGTGTGTTCAGAGGACGTGCTGCGGTTCCTCATCGCGAGGGCCT 1085
288 nHrArgGlnLeuaspValaIgluLeuArgGlyArgArgLeuArgA 305
1086 CCAGCGCGCTGGCAGCAGGACATGAGGTGCGAGCGGCGACATCCGCA 1135
305 rgsAspAspValaValaValaLeuAlaIgluAlaIAsnArgAspProArg 321
1136 AGGGAGAGGGCGTGTCTTCTCGACCTCGCTGATCAACCGCGACCGCGAC 1185
322 ArgTyRasPProAspAspAspIleuArgAspProValProSe 338
1186 GTGTTCCTCCCGGGCGGAGACATCGACTGGAGCCCGCCCGCCCATCA 1235
338 rMetSerPheGlyAlaIgluMetArgTyRcysLeuGlySerTyRleuAla 355
1236 CCTCGCCTTGGCTTCGAGTCCAGCCAGTGCCTGGCGCAGAACCTTGCC 1285
355 rgrHrIleuArgAlaAlaValaIAlaLeu...AlaArgLeuProGly 370
1286 GCGCGGAGCTGACATGCGATGCGACCTGTTCGAGCGGCTTCGCGG 1335
371 leuArgIleuGlycysAlaSerAspAlaLeuAlaTyRgluProArgHrme 387
1336 CTCAGGCTCGCGTACCCGCGCAGAGATCCGTCACAAAGCGGGGAGCAC 1385
387 tPheArgIleuAlaSerLeuProIleAlaPhe 398
1386 GATCCAGGGCTCTCTCGACCTGCGCGTGGCTGG 1419

```

seq_name: /cgn2_6/ptodata/1/lna/PCTUS_COMB.seq:PCT-US92-10885-10

seq_documentation_block:

```

; Sequence 10. Application PC/TUS9210885
; GENERAL INFORMATION:
; APPLICANT: SARIASIANI, SIMA
; TITLE OF INVENTION: CONSTITUTIVE
; TITLE OF INVENTION: EXPRESSION OF P450SOY
; TITLE OF INVENTION: AND FERREDOXIN-SOY IN
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: E. I. DU PONT DE MEMOURS
; ADDRESSER: AND COMPANY
; STREET: 1007 MARKET STREET
; CITY: WILMINGTON
; STATE: DELAWARE
; COUNTRY: USA
; ZIP: 19898
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch,
; MEDIUM TYPE: 1.0 MB
; COMPUTER: Macintosh
; OPERATING SYSTEM: Macintosh System, 6.0
; SOFTWARE: Microsoft Word, 4.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US92/10885
; FILING DATE: 19921216
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: GALLEGOS, R. THOMAS
; REGISTRATION NUMBER: 32,692
; REFERENCE/DOCKET NUMBER: CR-9000-A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 302-892-7342
; TELEFAX: 302-892-7949
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1735 base pairs

```

```

; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; PCT-US92-10885-10

```

```

alignment_scores:
  Quality: 410.00      Length: 428
  Ratio: 1.760        Gaps: 10
  Percent Similarity: 54.439  Percent Identity: 29.439

```

alignment_block:

US-09-724-797-36 x PCT-US92-10885-10 ..

Align seg 1/1 to: PCT-US92-10885-10 from: 1 to: 1735

```

10 AspProThrAspAlaAspValaIgluArgAspProTyRProSer..... 23
223 GACCCACAC.....TCCCGCGCCCGCGGAGCTCCTT 254
24 .....TyrHisTrpLeuAlaRgHisAspProV 33
255 CCCGAGAGACCGGGGTGCCCTTACACCCCGCCGCGGTACGACCCG 304
33 AlHisArgGly.....AlaHisArg 39
305 TCCGCGAGGCGCGCGCGGTGACCGGGTACCCCTTCGACGAGAGCCG 354
40 ValTrpTyRValSerArgPheAlaAspValaIgluAlaValIleuGlyAsp.. 55
355 GTCTGGGCGTACCGGCGACCGCTGGCGCTGCTGCTGCTGCTGCTGCTG 404
56 .....GluA 57
405 GCGGCTCTCCACCGACCGGACCGACCCGACCTCCCGCGCGGCGAGC 454
57 rgrPheAlaArgThrGlyIleArgArgPheTrpHrAspLeuValIgluPro 73
455 GGTTCGCGCGCGCGGACGCGCGCGCGCTC..... 483
74 GlyLeuLeuAlaGluIleValaIgluAspIleIleLeuPheGlnAspGluPr 90
484 .....GCTTGTCTCGCGGTGACGAGACC 506
90 oAspHisGlyArgLeuArgGlyValaIgluProAlaPheSerProSerA 107
507 CGAGCACAACACCCAGCGCGAGTGTCTATCCGACCTTCGTGTAAGC 556
107 leuAlaArgArgLeuGluProValIleAlaGlyThrValaIAspAspLeu 123
557 GGATCGCGCGCTCGCGCGGTATCCAGAGACCGTGAACGCGCTCTC 606
124 ArgProAlaLeuAlaArgGly...AlaMetAspValaIAspIleuAla 139
607 GACCGCATGAGCAGACAGAGGCGCGCGCGGACATGTAAGCGCGCTTCC 656
139 aTyRProLeuAlaLeuArgAlaValIleuGlyLeuGlyLeuProAlaA 156
657 CTGCGCGGTCCGCTCGATGTGTCTGTCTGCTGCTGCGGCGTCCCTAG 706
156 laAspTrpGlyAlaValaIgluArgTrpSerArgAspValaIgluArgHrleu 172
707 CCGACACACGCG.....TTCTTCGAGGAGACGCTCGCAGCAGCTC 744
173 AspArgGlyAlaSerAlaIgluAspMetArgArgGlyHisAlaIleAla 189
745 CTGCGCGCGCGCGGAGCGCAGATGTGAACAG.....GCCGCGCA 785
189 agluPheAlaAspTyRValaIgluArgAlaLeuAlaIgluArgArgGlu 206
786 CGAATCTGAGGAGTACTGCGCGCGCTGATCGACCGCAAGAGGGCGAGC 835

```

```

206 1yG1yG1uAspLeu...AlaLeuMetLeuAspAlaHisAspArgGly 221
      |||::: |||::: |||::: |||
836 CGGGTACGCGCTCTGTGACGAGCTGATCCACGGGACACCGGACGCA 885
222 LeuMetSerArgAsnGluLeuValSerThrValValThrPheIlePhe 238
      ::::: |||::: |||::: |||::: |||::: |||::: |||::: |||
886 CGGGTACGCGCGACGAGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCT 935
238 rG1yH1SG1uThrValAlaSerGlnValG1yAsnAlaValLeuSerLeu 255
      |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||
936 CGGGACAGAGACGAGCGACATGATCTGCTGCGGACGCTGACGCTGCG 985
255 euAlaHisProAspGlnLeuAspLeuValArgArgProAspLeu 271
      |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||
986 TGAGCCACCCCGAACAGCTGGCGGCTGCGGCGGCGGCGGACGACCC 1035
272 AlAG1AlaValG1uG1yG1yAspArgTyraProSerValG1SerAs 288
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1036 GCGGTGTGTGTGAGAGAGCTGCTGCGGTCTCTCATCGCGGAGGCT 1085
288 nThrArgGlnLeuAspValAspValG1uLeuArgG1yArgArgLeuArg 305
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1086 CCAGCGCTGTGCGACCGAGCATGAGGTGCGAGCGGCGGACGATCCCA 1135
305 rGAspAspValValValValLeuAlaG1yAlaAlaAsnArgAspProArg 321
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1136 AGGGGAGGCGGTGTGTCTTCTGACCTGCTGATCAACCGCGACCGCNC 1185
322 ArgTyraAspArgProAspPheAspIleG1yArgAspProValProSe 338
      ::: ||| ::: ||| ::: ||| ::: ||| ::: ||| ::: |||
1186 GTGTTCCTCCCGCGGCGAGACATCGACTGGACCGCGCGCGCGCATCA 1235
338 MetSerPheG1yAlaG1yMetArgTyraSerLeuG1ySerTyraAla 355
      |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||
1236 CCGTGCCTTGCTGTGAGTCCACCACTGCGGCGGCGGACGATCGCC 1285
355 rG1yH1SG1uLeuArgAlaAlaValAlaLeu...AlaArgLeuProGly 370
      |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||
1286 GCGCGGAGCTGACATCGGATCGGATCGGATCGGATCGGATCGGATCG 1335
371 LeuArgLeuG1yCysAlaSerAspAlaLeuAlaTyraGlnProArgThi 387
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1336 CTGAGCTGCGCGTACCGCGGCGAGATCGCTGACAAAGCGGCGGACAC 1385
387 tPheArgG1yLeuAlaSerLeuProIleAlaPhe 398
      ::::: |||::: |||::: |||::: |||::: |||::: |||::: |||
1386 GATCCAGGCGCTCTGACCTGCGCGCTGCGCTGCG 1419
seq_name: /cgn2.6/ptodata/1/ina/6B_COMB.seq:US-09-443-501A-2
seq_documentation_block:
; Sequence 2, Application US/09443501A
; Patent No. 6303342
; GENERAL INFORMATION:
; APPLICANT: Kusan Biosciences, Inc.
; APPLICANT: Uilleam, Bryan
; APPLICANT: Katz, Leonard
; APPLICANT: Khosla, Chaitan
; APPLICANT: Tang, Li
; APPLICANT: Ziermann, Rainer
; TITLE OF INVENTION: Recombinant Methods and Materials for Producing
; TITLE OF INVENTION: Epothilone and Epothilone Derivatives
; FILE REFERENCE: 30062-20031.00
; CURRENT APPLICATION NUMBER: US/09/443, 501A
; PRIOR APPLICATION NUMBER: US 60/130,560
; PRIOR FILING DATE: 1999-04-22
; PRIOR APPLICATION NUMBER: US 60/122,620
; PRIOR FILING DATE: 1999-03-03
; PRIOR APPLICATION NUMBER: US 60/119,386
; PRIOR FILING DATE: 1999-02-10
; PRIOR APPLICATION NUMBER: US 60/109,401
; PRIOR FILING DATE: 1998-11-20

```

```

; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 71989
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic construct
US-09-443-501A-2

```

```

alignment_scores:
  Quality: 402.50      Length: 404
  Ratio: 1.656
  Percent Similarity: 60.149      Percent Identity: 31.188

```

```

alignment_block:
US-09-724-797-36 x US-09-443-501A-2

```

```

Align seq 1/1 to: US-09-443-501A-2 from: 1 to: 71989

```

```

9 PheAspProThrAspAlaAspValArgArgAspProTyraProSerTyra 25
  ||| |||
56805 TTCAGCGCTTCCGCGCTGAGTACGCGGAGACCCGCTTCCGCGATCGA 56854
25 strLeuLeuArgHisAspProVal...HisArgG1yAlaHisArgValT 41
  ||| |||
56855 GCGCCTGAGAGAGCAACCCCATCTTCTTACGTGGATGAAAGCGCGCTCT 56904
41 rP1yValSerArgPheAlaAspValArgAlaValLeuG1yAspG1yArg 57
  ||| |||
56905 GGCTCCTCAACCGATCCACAGCGTCCGCGGTGCTGCGGACGACGACG 56954
58 PheAlaArgThrG1yLeuArgArgPheThrP1yAspLeuValG1yProG1 74
  ||| |||
56955 TTCGCGGTAGT...CGAGAAAGATGGAATGAGCGCGGAGTACTG 56998
74 yLeuLeuAlaG1uIleValG1yAspIleLeu...LeuPheGlnA 88
  ||| |||
56999 GTGCGGATCCCGGACGCTCAGGATGTAAGAAATGACGATGTTCGGGC 57048
88 spG1uPro...AspHisG1yArgLeuArgG1yValValG1yProAlaPhe 103
  ||| |||
57049 TGCGCGCGGAGATCAGCTGCGGTCCGCAAGAGTTCGATCCGATCTT 57098
104 SerProSerAlaLeuArgArgLeuG1uProValIleAlaG1yThrValAs 120
  ||| |||
57099 ACCTCAGCGCGATTCGACCTGCTGCGCGCGCAATGACGCGACGCTGA 57148
120 pasPLeuLeuArgProAlaLeuAlaArgG1yAlaMetAspValValAsp 137
  ||| |||
57149 CCAGCTGCTGATGCTGCTGCTCCGACAAAGAGAGTTCGATGTCGCG 57198
137 IuLeuAlaTyraProLeuAlaLeuArgAlaValLeuG1yLeuLeuG1y 153
  ||| |||
57199 ATTACGCGAGGAGATCCGATGCTGCGATGACGCTCTGTTGAAGTT 57248
154 ProAlaAlaAspTrpG1yAlaValG1yArgTrpSerArgAspValG1yAr 170
  ||| |||
57249 CCGGCGAGGTGACGAGAAAGTTCGCTGCTGCGTCCGCGACGCGCG 57298
170 g1yH1SG1uAspArgG1yAlaSerAla...G1uAspMetArgAspG 184
  ||| |||
57299 CGGCTGCGCGGTGTGTGCTGCGCGGCTCGATGAGGAGCAACGACCC 57348
184 IyH1SG1uAlaAlaIleAlaG1uPheAlaAspTyraValG1uArgAlaAla 200
  ||| |||
57349 TGATCGGCTCGTCCAGGAGGCGCTGCGCTGCTCATGCGCTCTCAT 57398
201 ArgArgArgG1uG1yG1yG1y...AspLeuLeuAlaLeuMetLeuAs 216
  ||| |||
57399 GAGCGCGCGAGAACCCGCTCGAAATGACGCTTGCAGATGCTGCTTCA 57448

```


OM of: US-09-724-797-36 to: EST:* out-format : pfs
Date: Jun 21, 2002 5:40 PM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

-MODEL=frame-p2n.model -DEV=xlh
-O=/cgn2.1/USPTO.spool/US09724797/runat_21062002.092538.18297/app-query.fasta.1.466
-DB=EST -OPMT=fastap -SUFFIX=1st -GAPOP=12.000 -GAPEXT=4.000
-MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000 -GAPOP=4.500
-OCAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000
-FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELXT=6.000
-DELXT=7.000 -STARP=1 -MATRIX=blotsum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0
-ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NOR=ext -HEAPSIZE=500
-MINLEN=0 -MAXLEN=200000000 -USER=US09724797.@CGN1.1.6398
-NCPU=6 -ICPU=3 -LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30
-NO_XLIPXY -WAIT -THREADS=1

Search information block:

Query: US-09-724-797-36
Query length: 402
Database: EST:*
Database sequences: 13736207
Database length: -1841457050
Search time (sec): 1687.050000

score list:

Sequence	Strd	Orig	ZScore	EScore	Len	Documentation
gb_hlc:AK004861	+	203.00	313.09	3.4e-08	3004	AK004861 Mus musculus adult ma
gb_est1:AA986024	+	187.00	300.27	1.8e-07	768	AA986024 uc72b04.y1 Sugano mus
gb_est1:AI530226	+	184.00	295.36	3.3e-07	762	AI530226 u190e12.y1 Sugano mus
gb_hlc:BC001776	+	182.50	280.94	2.1e-06	2492	BC001776 Homo sapiens, similar
gb_hlc:BC020267	+	182.50	280.94	2.1e-06	2492	BC020267 Homo sapiens, similar
gb_est1:BI246899	+	182.00	292.62	4.7e-07	717	BI246899 602959455f1 NCI, CGAP-I
gb_est1:BG204536	+	179.50	287.54	9.1e-07	789	BG204536 RST73942 Athersys RAGE
gb_est1:BG206832	+	179.00	287.15	9.5e-07	755	BG206832 RST6133 Athersys RAGE
gb_est1:BG182832	+	179.00	287.02	9.7e-07	765	BG182832 RST1693 Athersys RAGE
gb_est1:BI319542	+	177.50	283.91	1.4e-06	814	BI319542 m187b03.y1 Soares mus
gb_hlc:AK004724	+	176.50	282.07	1.8e-06	828	AK004724 RST15025 Athersys RAGE
gb_hlc:AK004724	+	176.00	268.51	1.0e-05	2930	AK004724 Mus musculus adult ma
gb_est1:BG189682	+	175.00	279.72	2.5e-06	817	BG189682 RST8722 Athersys RAGE
gb_est1:BG208878	+	174.00	278.17	3.0e-06	808	BG208878 RST8269 Athersys RAGE
gb_est1:BG197314	+	173.00	276.57	3.7e-06	803	BG197314 RST16556 Athersys RAGE
gb_hlc:AK007863	+	172.00	267.33	1.2e-05	1704	AK007863 Mus musculus 10 day c
gb_est1:AM941785	+	171.00	272.34	6.4e-06	888	AM941785 GH05994.3p1ime GH Dros
gb_gss:BH614466	+	170.50	274.35	4.9e-06	663	BH614466 2B8.Cont1122 Subloner
gb_gss:BH614466	+	170.50	274.35	4.9e-06	663	BH614466 2B8.Cont1122 Subloner
gb_est1:BG189683	+	169.50	271.23	7.3e-06	832	BG189683 RST73942 Athersys RAGE
gb_gss:AZ934336	+	169.00	273.07	5.8e-06	578	AZ934336 BJ_Ba0002013r B. japo
gb_gss:AZ934336	+	169.00	273.07	5.8e-06	578	AZ934336 BJ_Ba0002013r B. japo
gb_est1:BG110225	+	168.50	272.89	5.9e-06	551	BG110225 RST22387 Athersys RAGE
gb_est1:BG203018	+	167.00	266.51	1.3e-05	821	BG203018 RST19278 Athersys RAGE
gb_est1:BG203018	+	167.00	266.51	1.3e-05	821	BG203018 RST19278 Athersys RAGE
gb_est1:BG199924	+	166.50	265.62	1.5e-05	816	BG199924 RST19220 Athersys RAGE
gb_est1:BG133781	+	166.00	265.87	1.5e-05	816	BG133781 RST19220 Athersys RAGE
gb_hlc:AK002588	+	165.50	265.87	1.5e-05	733	AK002588 Mus musculus adult ma
gb_est1:BG202472	+	165.00	263.09	2.4e-05	824	BG202472 RST18259 Athersys RAGE
gb_est1:BG141013	+	164.00	262.03	2.4e-05	767	BG141013 RST18259 Athersys RAGE
gb_est1:BG206832	+	164.00	261.79	2.5e-05	790	BG206832 RST26134 Athersys RAGE
gb_est1:BG194899	+	164.00	261.65	2.5e-05	801	BG194899 RST13956 Athersys RAGE
gb_est1:BG12937	+	163.50	261.19	2.7e-05	772	BG12937 602983306f1 NCI, CGAP-I
gb_est1:BI247661	+	163.00	263.26	2.0e-05	579	BI247661 602959455f1 NCI, CGAP-I
gb_est1:BI245034	+	163.00	259.82	3.2e-05	815	BI245034 602959455f1 NCI, CGAP-I
gb_est1:BG207705	+	162.00	259.13	3.5e-05	740	BG207705 RST27188 Athersys RAGE
gb_est1:BG191712	+	162.00	258.78	3.6e-05	766	BG191712 RST10808 Athersys RAGE
gb_est1:BG136771	+	160.00	256.59	4.8e-05	685	BG136771 RST136271 unpublished c
gb_est1:BI150948	+	160.00	256.03	5.2e-05	724	BI150948 BI150948 unpublished c
gb_est1:AL564426	-	158.00	250.63	0.0001	890	AL564426 AL564426 LTI_NFY001_NH

gb_est1:BG185434 + 156.50 248.17 0.0001 887 | BG185434 RST4377 Athersys R
gb_est1:BG292017 + 156.00 249.89 0.0001 689 | BG292017 HNC5-1-F9.R HNC (H
gb_est1:BG200435 + 155.00 246.69 0.0002 802 | BG200435 RST19748 Athersys
gb_est1:BG205065 + 153.00 243.21 0.0003 815 | BG205065 RST24484 Athersys
gb_est1:AW257685 + 150.50 241.17 0.0003 661 | AW257685 SWD25CA00808SK B.

seq_name: gb_hlc:AK004861

seq_documentation_block:

LOCUS AK004861 3004 bp mRNA linear HTC 19-JAN-2002
DEFINITION Mus musculus adult male liver cDNA, RIKEN full-length enriched
library, clone:1300002013:cytoschrome p450, steroid inducible 3a11,
full insert sequence.

ACCESSION AK004861.1 GI:12836364

VERSION AK004861.1

KEYWORDS HTC; CAP trapper.

SOURCE Mus musculus (strain:C57BL/6J) adult male liver cDNA to mRNA,
clone:1300002013.

ORGANISM

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (sites)

Carninci, P. and Hayashizaki, Y.

High-efficiency full-length cDNA cloning

99279253

10349636

2 (sites)

Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.

Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes

Genome Res. 10 (10), 1617-1630 (2000)

20499374

11042159

3 (sites)

Shibata, K., Itoh, M., Aizawa, K., Nagao, S., Sasaki, N., Carninci, P.,
Kono, H., Akiyama, J., Nishi, K., Katsunari, T., Tashiro, H., Itoh, M.,
Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishino, T., Harada, A.,
Yamamoto, R., Matsunoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,
Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watanabe, M.,
Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsunaga, S., Kawai, J.,
Okazaki, Y., Muramatsu, M., Inoue, Y., Kita, A. and Hayashizaki, Y.

RIKEN integrated sequence analysis (RISA) system--384-format

Sequencing pipeline with 384 multipillar sequencer

Genome Res. 10 (11), 1757-1771 (2000)

20530913

11076861

4 (sites)

The RIKEN Genome Exploration Research Group Phase II Team and the
PANTOM Consortium.

Functional annotation of a full-length mouse cDNA collection

Nature 409, 685-690 (2001)

5 (bases 1 to 3004)

Arakawa, T., Balderelli, R., Bono, H., Brownstein, M., Bull, C.,
Carninci, P., Fukuda, S., Fukunishi, Y., Furuno, M., Hanagaki, T.,
Hara, A., Hayatsu, N., Hill, D., Hiramoto, K., Hiraoka, T., Hori, F.,
Hume, D., Imotani, K., Ishii, Y., Itoh, M., Izawa, M., Katsunari, T.,
Kato, H., Kawai, J., Kojima, Y., Kono, H., Kouda, M., Koyama, S.,
Kurihara, C., Matsuyama, T., Miyazaki, A., Nishi, K., Nomura, K.,
Numazaki, R., Ohno, M., Okazaki, Y., Okido, T., Owa, C., Quackenbush, J.,
Saito, H., Saito, R., Sakai, C., Sakai, K., Sano, H., Saeki, D.,
Schiraldi, L., Shibata, K., Shibata, Y., Shingawa, A., Shiraki, T.,
Sogabe, Y., Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F.,
Tanaka, T., Tejima, Y., Toyota, T., Yamamura, T., Yamakawa, I.,
Yasunishi, A., Yoshida, K., Yoshino, M., Muramatsu, M. and
Hayashizaki, Y.

Direct Submission

Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of
Physical and Chemical Research (RIKEN), Laboratory for Genome
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),

[illegible]

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. consortium/INL at: <http://image.lnl.gov>
Series: IRAL Plate: 6 Row: h Column: 9
This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 13904864
This clone has the following problem: frame shifted.

/db_xref="taxon:9606"
/clone="IMAGE:3354171"
/issue_type="Eye, retinoblastoma"
/clone_lib="NIH MGCL6"
/lab_host="DH10B-R"
/note="Vector: POTB7"
BASE COUNT 512 a 727 c 679 g 574 t
ORIGIN

alignment_scores:

Quality: 182.50 Length: 486
Ratio: 0.882 Gaps: 21
Percent Similarity: 42.593 Percent Identity: 23.045

alignment_block:

US-09-724-797-36 x BC001776 ..

Align seg 1/1 to: BC001776 from: 1 to: 2492

```

14 AlaAspValArgAspProTyrProSerTyr..... 24
16 GCAGACATCCAGGCCCTCAGCCGCTTTCTGCGCCGAACCTTCTG 215
25 .....HistripleuLeuArgHisAspProValH 34
216 CAAGGGGGGCTGTCAGAGCTACAGAGTGTGAGGCGCGCGCGC 265
34 IsArgLYAlaHisArgValTrrpTyrValSerArgPheAlaAspValArg 50
266 ACTTCGGGCGCG.....GTGTGCTAGCCAGC...TTTGGAGAGTGGC 306
51 AlaVal.....LeuLYAspGluArgPheAlaArgTth 61
307 ACCGTGTACGTGTGCTGCCCCCTGACACTGTGTCAGAGAGCTGTCGACAGGA 356
61 rGlyLeuArg.....ArgPheTrrpThrasp..... 69
357 GGGACCCCGGCGCGAGCGCTGACACTTCTCCCTCGAGCGAGGACCGCC 406
70 .....LeuValGlyProGlyLeuAlaGluLeuValGly 81
407 GCTGCCCGCAGCGGCTTGGCGGA..... 429
82 AspIleLeuPheGlnAspGluProAspHisGlyArgLeuArgGlyVal 98
430 .....CTGCTCACTGCGGAGGAGGAGAAATGCAAGGCTCCCGCAGTCT 473
98 lValGlyPro...AlaPheSerProSerAlaLeuArgArgLeuGluProv 114
474 CTTGCCCCCTGCTCCTCCCTCAAGCGCGCGCGCTACGCGGAA 523
114 allIleAlaGlyThrValAspAspleuLeuArg..... 124
524 CCCTGACACAGTATGCTGCGACTGTGCGGCTGTGAGGCGCGCAGCG 573
125 .....ProAlaLeuAlaArgGlyAlaMetAspValVal 135
574 GGACGTGCGACAGGCGCGCGCGCTGTGG.....GACGTGGC 614
135 lAspGlyLeu..... 138
615 GGGGGAATTTCACAAATTTCGAGCTGGAAGTGAATCCAGACAGAGCTG 664
139 .....AlaTyr 140
665 GGCAGGCGTGGGGGCGCTACCAAGAGCTCCGGAAACCTGACGCGCG 714
141 ProLeuAlaLeuArgAlaValLeuGly.....LeuLeuGlyLe 153
715 CCCTCCCGCAAGAGCATGCGCGGCTTCTGCTGCTGCGCTCCGCGCTTGGGCT 764
153 uProAlaAlaAspTrrpGlyAlaValGlyArgTrrpSerArgAspValGly 170

```

```

765 GCGTGA.....GGCTCAAGTGGCACCCGACAGCGAGACTTCATC 805
170 rGthrLeuAspArgGlyAlaSerAlaGluAspMetArgArgGlyHisAla 186
806 CCGCTGTGGGCTCGGTGTTGTGTCACGCTTGAACATGAGCATGCC 855
187 AlaIleAla..... 189
856 CCACTGGCTGGCCACCTTGTGCTGGGCGCTGGGCGCTCTGCGGAG 905
190 .....Glu.PheAlaAspTyrValGluArgAlaLeuAlaArgArg 202
906 ACTGGAGCAGATGTTGCA...TTTGTCTAGAGGACAGTGGAGCGGCGA 952
203 Arg.....ArgGluGlyGlyGluAspLeuAlaLeuMet 214
953 GAGGACAGAGGACCCATGAGAGGAGGAGGACAGCCCGAGAGAGCTTGA 1002
214 tLeuAspAlaHisAspArgGlyLeuMetSerArgAsnGlu..... 227
1003 GTCTGGGGCGCACCTGACCCACTTCTCTCCGGAGAGATTGCTGCC 1052
228 .....IleValSerThrValValThrPheIlePheThrGlyHisGluThr 242
1053 AGTCATCTCGTGGAAATGTACAGAGATTGCTATTGGCGGAGTGACAGC 1102
243 ValAlaSerGlnValGlyAsnAlaValLeuSerLeuLeuAlaHisProAs 259
1103 GTGTCCAAACGCTCTGTGGGCTGTGTATGAGCTCTCCCGGACCCGGA 1152
259 P..... 259
1153 AGTCCAGACAGCACTCCACTCAGAGATCACAGCTGCCCTGAGCGCTGCT 1202
260 .....GlnLeuAspLeuLeuArgArgArgProAspleuLeu 271
1203 CCACTGCTACCCCTCAGCCACTGCTGTCTGCCAGCTGCC...CTGCTG 1249
272 AlaGlnAlaValGluGlyCysLeuArgTyrAspProSerValGlnSerAs 288
1250 AAGCGGTGTCAAGAGAGTGTAAAGACTGTACCTGCTGTGCTGAGAA 1299
288 nThrArgGlnLeuAspValAspValGluLeuArgGlyArgArgLeuArg 305
1300 TTCTCGTGTCCACAGACAAACATTCATGTGGTGTCTATATTATCCCA 1349
305 rGAspAspValValValLeuAlaGlyAlaAlaAsnArgAspProArg 321
1350 AAATACGCTGTGCTACTGTGTGCTACTATAGCCACTTCAAGGAGCCCTGCC 1399
322 ArgTyrAspArgProAspAspPheAspIleGluArg..... 333
1400 CAATTCACAGAGCCAAATTCCTTCTGTCAGCTCCGCTGGCGGAGGG 1449
334 .....AspProValProSerMetSerPheGlyAlaGlyMetArgT 347
1450 TCCACACCCCGACACCATTTGCATCTCTTCCCTTTGGCTTTGGCAGCGCA 1499
347 yTCysLeuGlySerTyrLeuAlaArgThrGlnLeuArgAlaAlaValAla 363
1500 GCTGTATGGGAGAGACCGCTGGCAGAGCTTGAATTGCAATTTGGCTTTGGCC 1549
364 AlaLeu 365
1550 CAGATC 1555

```

seq_name: gb_hlc:BC020267

seq_documentation_block: 2492 bp mRNA linear HTC 19-DEC-2001
LOCUS BC020267
DEFINITION Homo sapiens, clone IMAGE:4580963, mRNA.
ACCESSION BC020267

VERSION	BC020267.1	GI:17999623
KEYWORDS	HTC.	
SOURCE	human.	
ORGANISM	Homo sapiens	
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.	
AUTHORS	1 (bases 1 to 2492)	
TITLE	Strausberg,R.	
JOURNAL	Direct Submission	
	Submitted (19-DEC-2001) National Institutes of Health, Mammalian	

REMARK	COMMENT
NIH-MGC Project URL: http://mgc.ncl.nih.gov	Contact: MGC help desk

BC Cancer Agency, Vancouver, BC, Canada
info@bcgsc.bc.ca

Steven Jones, Jennifer Asano, Ian Bosdel, Aaron Butterfield, Susanna Chan, Readman Chiu, Chris Fell, Erin Garland, Ran Guin, Leticia Hsiao, Martin Krzywniaki, Reta Kutscher, Oliver Lee, Soo Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeay, Steven Ness, Pawan Pandoh, Anna-Liisa Prabh, Parvaneh Speedi, Jacqueline Schein, Duane Smallus, Michael Smith, Lorraine Spence, Jeff Stolt, Michael Thorne, Miranada Tsai, Natasja Van den Bosch, Jill Vardy, George Yang, Scott Zuyderduyn, Marco Zatta.

This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA g1: 13904864 This clone has the following problem: no cloning site / microdeletion.

FEATURES	Location/Qualifiers
source	1. .2492

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4580963"
/tissue_type="Eye, retinoblastoma"
/clone_lib="NRH MGC_16"
/lab_host="DH10B-R"
/notes="Vector: pOTB7"

```

alignment_scores:		
Quality:	182.50	Length: 486
Ratio:	0.882	Gaps: 21
Percent Similarity:	42.593	Percent Identity: 23.045

```
alignment_block:
```

Align seg 1/1 to: BC020267 from: 1 to: 2492

14 AlaaspValAlaArgAspProIyrProSerIyr..... 24
 |||||::: |||||:::
 16 GCAGACATCCAGAGCCCTCTACGCCAGCTTTCTGGCCGAATTTTGTG 215
 25 HistPleuLeuArgHisAspProValH 34
 216 CAAGGGGGGGCTGTGAGGCTACAGAGCTCAGAGTCAGAGGCCGCCCGC 265
 34 IsArgIyValaHHisArgValITrPyrValISerArgPheAlaAspValArg 50
 || ||| |||||::: |||||:::
 266 ACTTCGCGGCCG..... GGTGGCTAGCCACAC.....TTTGGACAGTGC 306

51 AlaVala.....LeuAlaYspGIuArgPheAlaArgTh 61
 307 ACCGGTACGTGGCTGGCCCTGTCACTCTGTGAGAGCTCTGCACAG 356
 61 rGlyIleArg.....ArgPheTrpIAsp..... 69
 357 GGGACCCCGGCGCCGAGCGCTGACCTTCTCCCGGAGGAGACCGCC 406
 70LeuValGlyProGlyLeuLeuAlaGluIleValGly 81
 407 GCTGCCGCCAGCGGGCTTCCGA..... 429
 82 AspIleIleLeuPheGluAspIuProAspHisGlyArgLeuArgGlyVa 98
 430CTGGCTACGTGGGAGGAGCAAGATGGCAAGAGCTCCGACGCT 473
 98 ValGlyPro...AlaPheSerProSerAlaLeuArgArgLeuIuProV 114
 474 CCGGGCCCCGCTCTCTCTCCGGCTCAAGCGGCGCCCGCTACCGCGGA 523
 114 alIleAlaGlyThrValAspAspLeuLeuArg..... 124
 524 CCTGTACACAGCTACTCTGTGCACCTTGGCGCGCTGAGGCGCCAGCG 573
 125ProAlaLeuAlaArgGlyAlaMetAspValVa 135
 574 GAGACGTGGACGCGGGCCCGCCCGCTGTGTGG.....GACGTGGC 614
 135 AspIuLeu..... 138
 615 GGGGGAATTTTACAGTTCCGACTGGAAAGTGAGTCCAGACAGAGCTG 664
 139AlaTyr 140
 665 GGCAGGCGCGGGGGCGGCTTACAGACCTTCCCGAACCCTGACGGCGC 714
 141 ProIeuAlaLeuArgAlaValLeuGly.....LeuLeuGlyLe 153
 715 CCCCTCCGACAGACAGCATCGCCGCGGTCTGTCTGCGCTGGCTTGGCT 764
 153 uProAlaAlaAspTrpGlyAlaValGlyArgTrpSerArgAspValGly 170
 765 GCCTGGA.....GCTCAAGTGCACCCGACACGAGAACCTTCATC 805
 170 rGThIleuAspArgGlyAlaSerAlaGluAspMetArgArgGlyHisAla 186
 806 CGCGCTGGGCTCGGTGTGTGTGTCTCAAGCTGTGACCAATGGCGATCC 855
 187 AlaIleAla..... 189
 856 CCACGTGGCTGGCCCACTTGTCCCGGGGCCCGGGGGCGGCTGGCGAG 905
 190Glu.PheAlaAspTrValGluArgAlaLeuAlaArgArg 202
 906 ACTGGGACCGAGATGTTTGA...TTTGTCAAGGCAAGCTGAGACGGGA 952
 203 Arg.....ArgGluGlyGluAspLeuLeuAlaLeuMe 214
 953 GAGGCGAGAGCCCATGAGACGAGAGGAGACGCCCCGAGAAAGACTGGA 1002
 214 tLeuAspAlaHisAspArgGlyLeuMetSerArgAsnGlu..... 227
 1003 GCTGGGGGCGACCTGACCAACTTCTGTCCCGGAAAGATGCTCGTCCC 1052
 228IleValSerThrValValTrpPheIlePheThrGlyHisGluThr 242
 1053 AGTCATCTCTGGGAATGTGACAGAGTTGTCAATTGGCGGAGTGGACACG 1102
 243 ValAlaSerGluValGlyAsnAlaValLeuSerLeuAlaHisProAs 259
 1103 GTGTCCAAACGCTCTCTTGGCTCTGTATGATAGCTCTCCGACACCCGA 1152
 259 P..... 259

```

1153 AGTCAGACAGACGCTCCACTCAGAGATCAGAGCTCCCTGAGCCCTGGCT 1202
260 .....GlnLeuAspLeuLeuArgArgProAspLeu 271
1203 CCAGTCCCTACCCCTCAGCCACTGTCTGTCCAGCTGCC...CTGCTG 1249
272 AlaGlnAlaValGluGluCysLeuArgTyrAspProSerValGlnSerAs 288
1250 AAGCGGTGTGTCAGAGAGTGTAGACTGTACCTGTGTGTAACCTGAGAA 1299
288 nThrArgGlnLeuAspValAspValGluLeuArgGlyArgArgLeuArg 305
1300 TTCTGTGTCCAGACAAAGACATTCATGTGGTGACTATATATATATCCCC 1349
305 rGAspAspValValValAlaValAlaGlyAlaAlaAsnArgAspProArg 321
1350 AAAATAGCGTGTGCTACTGTGTCACTATGCACTTCAAGGAGACCTGCC 1399
322 ArgTyrAspArgProAspAspPheAspIleGluArg..... 333
1400 CAGTTCCTCCAGAGCCAAATCTTTCTCCAGCTCCCTGGCTGGGGAGAG 1449
334 .....AspProValProSerMetSerPheGlyAlaGlyMetArgT 347
1450 TCCACACCCCAACCCATTGTCATCTCTCCCTTGGCTTGGCAGAGCCA 1499
347 yrcysLeuGlySerTyrLeuAlaArgThrGlnLeuArgAlaAlaValAla 363
1500 GCGTATGGGGAGACCGCTGCGAGAGCTGTAATTCGAATGGCTTTGGCC 1549
364 AlaLeu 365
1550 CAGATC 1555

seq_name: gb_est2:BI246899
seq_documentation_block:
LOCUS BI246899 717 bp mRNA linear EST 17-JUL-2001
DEFINITION 602959455F1 NCI_CGAP_L19 Mus musculus cDNA clone IMAGE:5125068 5',
ACCESSION BI246899
VERSION BI246899.1 GI:14791308
KEYWORDS mRNA sequence.
SOURCE EST.
ORGANISM Mus musculus
            house mouse.
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
            1 (bases 1 to 717)
            NIH-MGC http://mgc.nci.nih.gov/.
            National Institutes of Health, Mammalian Gene Collection (MGC)
            Unpublished (1999)
            Contact: Robert Strausberg, Ph.D.
            Email: cga@bbs-rcmail.nih.gov
            Tissue Procurement: Jeffrey E. Green, M.D.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LNL at:
            http://image.llnl.gov
            Plate: LHAM1305 row: n column: 13
            High quality sequence stop: 717.
            Location/Qualifiers
                1..717
                /organism="Mus musculus"
                /strain="FVB/N"
                /db_xref="taxon:10090"
                /clone="IMAGE:5125068"
                /clone_1id="NCI CGAP L19"
                /lab_host="DH10B (TI phage-resistant)"
                /note="Organ: liver; Vector: pCMV-SPORT6; Site_1: NotI;
                Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT."

```

```

Average insert size 1.9 kb. Constructed by Life
Technologies. Note: this is a NCI_CGAP Library."
BASE COUNT      212 a      175 c      150 g      180 t
ORIGIN
alignment_scores:
    Quality: 182.00      Length: 179
    Ratio: 1.784
    Percent Similarity: 56.983      Percent Identity: 29.050
alignment_block:
US-09-724-797-36 x BI246899 ..
Align seg 1/1 to: BI246899 from: 1 to: 717
209 AspleuLeuAlaLeuMetLeuAspAlaHisArgGly..... 221
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
61 GATTCTCTCAGCTGATGATGATGATTCATATATATCCAAAGCAAGT 110
222 .....LeuMetSerArgAsnGluIleValSerThrValAlaIhrp 235
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
111 CTCTCATAAAGCCCTTCTGACATGAGATCACACCCAGTCATATATCT 160
235 heilePheThrGlyHisGluThrValAlaSerGlnValGlyAsnAlaVal 251
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
161 TTATTTTGTGGTGTATGAACACACAGTACACTTCTCCACCTG 210
252 LeuSerLeuLeuAlaHisPro.....AspGlnLe 261
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
211 CATTCCTTGGCCACTCACCTGATATCCAGAAAAAATGACAGATGAGAT 260
261 uAsp.....LeuLeuArgArgArgPro..... 268
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
261 CGATGAGGCTCTGCCACACAGGACCTCCACGATGATGATCTGATGAG 310
269 .....AspleuLeuAlaGlnAlaValGluGluCysLeuArgTyrAspPro 283
|||  |||  |||  |||  |||  |||  |||  |||
311 AGATGGAATACCTGATATGATGCTTAATGAACCTCAGATTATATCC 360
284 SerValGlnSerAsnThrArgGlnLeuAspValAspValGluLeuArgG1 300
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
361 ATGGCTAATGACTTGAGAGAGCTGTAGAAAGAGTGAACCTCAAGG 410
300 yArgArgLeuArgArgAspAspValValValAlaValAlaAla 317
|||  |||  |||  |||  |||  |||  |||  |||
411 TGTGTATATCCCAAGGATCAAGATGATGATTCATCTTATGCTTTC 460
317 snArgAspProArgArgTyrAspArgProAspAspPheAspIleGluArg 333
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
461 ACCATGACCCACAGACGCTGACAGCTGAAGAAATCCAACTTCAAGG 510
334 .....AspProValProSerMetSerPhe 341
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
511 TTCAGCAAGAGAAACAAAGGACACATTCATTCATATATATCCCTT 560
341 eGlyAlaGlyMetArgTyrCysLeuGlySerTyrLeu 353
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
561 TGGGAATGGACCCAGGAACTGCTTGCAATGAGGTG 597

seq_name: gb_est2:BG204536
seq_documentation_block:
LOCUS BG204536 789 bp mRNA linear EST 21-APR-2001
DEFINITION RST73942 Athersys RAGE Library Homo sapiens cDNA, mRNA sequence.
ACCESSION BG204536
VERSION BG204536.1 GI:13726223
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
            1 (bases 1 to 789)

```

AUTHORS
Harrington, J. J., Sherf, B., Rundlett, S., Jackson, P. D., Perry, R., Cain, S., Leventhal, C., Thornton, M., Ramachandran, R., Whittington, J., Lerner, L., Costanzo, D., McElligott, K., Booser, S., Mays, R., Smith, E., Veloso, N., Klika, A., Hess, J., Cothren, K., Lo, K., Offenbacher, J., Danzig, J., and Ducar, M.

TITLE
Creation of genome-wide protein expression libraries using random activation of gene expression

JOURNAL
Nat. Biotechnol. 19 (5), 440-445 (2001)

MEDLINE
21227151

COMMENT
Contact: Scott J. Cain
Athersys, Inc.
3201 Carnegie Ave, Cleveland, OH 44115, USA
Tel: 216 431 9900
Fax: 216 361 9596
Email: sca@atersys.com
High quality sequence stop: 507.

FEATURES
source
1..789
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="Athersys RAGE Library"
/cell_line="HT1080"
/note="See 'Creation of Genome-wide Protein Expression Libraries using Random Activation of Gene Expression', Nature Biotechnology, in press. Note that even though the cell type indicated is HT1080, since a random activation method was used, these sequence tags are not necessarily expressed in HT1080 under normal circumstances."

BASE COUNT
244 a 176 c 160 g 209 t

ORIGIN

alignment_scores:
Quality: 179.50 Length: 220
Ratio: 1.561 Gaps: 5
Percent Similarity: 52.273 Percent Identity: 26.364

alignment_block:
US-09-724-797-36 x BG204536 ..

Align seg 1/1 to: BG204536 from: 1 to: 789

```

190 GtUphEaAspTYrValGluArgAlaLeuAlaArgArGArgGluG1 206
    ||| :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
60 GAAGTCAAAATTTTAAAGAAATCTGTAAAAAGATGAAGAAAGTCG 109
206 yGlyGlu.....AspLeuLeuAlaLeuMetLeuA 216
    ||| :|||||:|||||:|||||:|||||:|||||:|||||:|||||
110 CCTGGAAGATACACAAAGACGCGATGCTTCCTCAGCTGATGATTG 159
216 spAlaHAsp.....ArgGlyLeuMetSerArgAsn 226
    ||| :|||||:|||||:|||||:|||||:|||||:|||||:|||||
160 ACTCTCAAAATTCAAAAGAACTGAGTCCACAAAGCTTTCCGATCTG 209
227 GluLeuAlaSerThrValAlaThrPheIlePheThrGlyHisGluThrVa 243
    ||| :|||||:|||||:|||||:|||||:|||||:|||||:|||||
210 GAGCTCGGCGCCCAATCAATATCTTATTTTCTGCTGATGAACACAC 259
243 lAlaSerGlnValGlyAsnAlaValLeuSerLeuLeuAlaHisProAspG 260
    ||| :|||||:|||||:|||||:|||||:|||||:|||||:|||||
260 GAGCGTCTCTCTCCCTTCATTATGATGAACTGCCACTCACCCTGATG 309
260 lLeuAspLeuLeuAlaArgArGArgProAsp..... 269
    ||| :|||||:|||||:|||||:|||||:|||||:|||||:|||||
310 TCCACGACAAACTGCAGAGCAAAATGTATGCAAGTTTACCACCAATAGCA 359
270 .....LeuLeuAlaGlnAlaVala 275
    ||| :|||||:|||||:|||||:|||||:|||||:|||||:|||||
360 CCACCACCTATGATCTGCTACAGATGAGATCTTGACATGCTGCT 409
275 lGluGluLucySerLeuArgTyArgProSerValGlnSerAsnThrArgGlnL 292
    ||| :|||||:|||||:|||||:|||||:|||||:|||||:|||||
410 GAATGAACGCTCAGATTTATTCCTCAATGCTATGAGACTGAGAGGCTCT 459

```

292 euaSpValaSpValGluLeuArgGlyValArgLeuArgArgAspAspVal 308
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
460 GCAAAAAGATGTGAGATCATGGATGTCATTCCTCCAAAGGGCTGTG 509
309 ValValValLeuAlaGlyAlaAlaAsnArgAspProAlaArgTyArgPar 325
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
510 GTGATGATTCACAGTATGCTCTTCACCGTGACCCAAAGTACTGACAGA 559
325 gProAspAspPheAspIleGluArg..... 333
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
560 GCCTGAGAGAGTCTCTCTGAAAGATTCAGCAAGACAGACAGACACA 609
334 ..AspProValProSerMetSerPheGlyAlaGlyMetArgTyArgCysLeu 349
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
610 TAGATCCTTACATATACACACCCCTTTGGAAGTGCAGCAAACTGCATT 659
350 GlySerTyLeuAlaAlaArgThGlnLeuArgAlaAlaValAlaAlaLeuAl 366
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
660 GGCATGAGAGTGTGCTCTCATGACATGAA.....CTTGCTCTAT 700
366 aArgLeuPro 369
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
701 CAGAAGTCTT 710

seq_name: gb_est2:BG206682

seq_documentation_block:
LOCUS BG206682 755 bp mRNA linear EST 21-APR-2001
DEFINITION R5226133 Athersys RAGE Library Homo sapiens cDNA, mRNA sequence.
ACCESSION BG206682
VERSION BG206682.1 GI:13728369
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 755)
AUTHORS
Harrington, J. J., Sherf, B., Rundlett, S., Jackson, P. D., Perry, R., Cain, S., Leventhal, C., Thornton, M., Ramachandran, R., Whittington, J., Lerner, L., Costanzo, D., McElligott, K., Booser, S., Mays, R., Smith, E., Veloso, N., Klika, A., Hess, J., Cothren, K., Lo, K., Offenbacher, J., Danzig, J., and Ducar, M.
TITLE
Creation of genome-wide protein expression libraries using random activation of gene expression
JOURNAL
Nat. Biotechnol. 19 (5), 440-445 (2001)
MEDLINE
21227151
COMMENT
Contact: Scott J. Cain
Athersys, Inc.
3201 Carnegie Ave, Cleveland, OH 44115, USA
Tel: 216 431 9900
Fax: 216 361 9596
Email: sca@atersys.com
High quality sequence stop: 496.

FEATURES
source
1..755
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="Athersys RAGE Library"
/cell_line="HT1080"
/note="See 'Creation of Genome-wide Protein Expression Libraries using Random Activation of Gene Expression', Nature Biotechnology, in press. Note that even though the cell type indicated is HT1080, since a random activation method was used, these sequence tags are not necessarily expressed in HT1080 under normal circumstances."

BASE COUNT
231 a 171 c 153 g 200 t

ORIGIN

alignment_scores:
Quality: 179.00 Length: 213
Ratio: 1.584 Gaps: 4

Percent Similarity: 53.052 Percent Identity: 25.822

alignment_block:

US-09-724-797-36 x BG206682 ..

Align seg 1/1 to: BG206682 from: 1 to: 755

```

190 GIuphealaasptYrValGluarGalaleuAlaArgArgArgGluG1 206
    ||| ..... ||| .....
60 GAAGTTACAAATTTTAAAGAAATCTGTAAAGATGAAAGAAAGCG 109
206 yglYglu.....AspleuAlaleuMetleuA 216
    ||| ..... ||| .....
110 CCTCGAAGATACACAAAGACCGAGTGGATTCTCTGAGCTGATGATG 159
216 spAlaHisasp.....ArgGlyLeuMetSerArgasn 226
    ||| ..... ||| .....
160 ACTCTCAAAATTCAAAAGAACTGAGTCCCAAAAGCTCTGTCCGATCTG 209
227 GluIleValSerThrValValThrPheIlePheThrGlyHisGluThrVa 243
    ||| ..... ||| .....
210 GAGTCGCGGCCCAATCAATTATCTTTATTTTGTGCTGCTATGAACCCAC 259
243 lAlaSerGlnValGlyAsnAlaValLeuSerLeuLeuAlaHisProaspG 260
    ||| ..... ||| .....
260 GAGCAGTGTCTCTCTTCATTATGTAAGAACTGGCCACTCACCCCTGATG 309
260 lInleuAspleuLeuArgArgProasp..... 269
    ||| ..... ||| .....
310 TCCAGCAAAACTGCAGAGGAATGTGATGCACTTTTACCAATAAGGCA 359
270 .....LeuAlaGlnAlaVala 275
    ||| ..... ||| .....
360 CCACCCACTATGATACTGCTGACAGATGAGATCTTGACATGAGTGGT 409
275 lGluGlyCysLeuArgTyrAspProSerValGlnSerAsnThrArgGlnL 292
    ||| ..... ||| .....
410 GAATGAACGCTCAGATTATTCCTCAATGTGATGAGACTTGAGAGGGTCT 459
292 euAspValAspValGluLeuArgGlyArgLeuArgArgAspAspVal 308
    ||| ..... ||| .....
460 GCAAAAAGATGTGAGATCAATGAGATGTTCAATCCCAAGGGGTGTG 509
309 ValValValLeuAlaGlyAlaAlaAsnArgAspProArgTyrAspPar 325
    ||| ..... ||| .....
510 GTGATGATTCACAGCTATGCTCTCACCGCCAAAGTACTGACAGACA 559
325 gProAspAspPheAspIleGluArg..... 333
    ||| ..... ||| .....
560 GCCTGAGAAAGTCTCCCTGAAAGATTCACAGACAGACAGACAGACACA 609
334 ..AspProValProSerMetSerPheGlyAlaGlyMetArgTyrCysLeu 349
    ||| ..... ||| .....
610 TAGATCTCTACATATACACACCCCTTTGGAGTGGACCCAGAACCTGAT 659
350 GlySerTyrLeuAlaArgThrGlnLeuAlaAlaVal 362
    ||| ..... ||| .....
660 GGCATGAGGTTTGTCTCATGACATGATAAACTTGCTCTA 698
seq_name: gb_est2:BG182832

```

seq documentation_block:

```

LOCUS      BG182832              765 bp      mRNA      linear      EST 21-APR-2001
DEFINITION RST1693 Atherysys RAGE Library Homo sapiens cDNA, mRNA sequence.
ACCESSION  BG182832
VERSION    BG182832.1  GI:13704503
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE  1 (bases 1 to 765)
AUTHORS   Harrington,J.J., Sherif,B., Rundlett,S., Jackson,P.D., Perry,R.,

```

Cain,S., Leventhal,C., Thornton,M., Ramchandran,R., Whittington,J., Lerner,L., Costanzo,D., McElligott,K., Boozet,S., Mays,R., Smith,E., Veloso,N., Kikla,A., Hess,J., Colhoun,K., Lo,K., Offenbacher,J., Danzig,J. and Ducar,M.
Creation of genome-wide protein expression libraries using random activation of gene expression
Nat. Biotechnol. 19 (5), 440-445 (2001)
21227151

COMMENT

Contact: Scott J. Cain
Athersys, Inc.
3201 Carnegie Ave, Cleveland, OH 44115, USA
Tel: 216 431 9900
Fax: 216 361 9596
Email: scain@atersys.com

FEATURES

High quality sequence stop: 506.
Location/Qualifiers
1..765
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="Athersys RAGE Library"
/cell_line="HT1080"
/note="See 'Creation of Genome-wide Protein Expression Libraries using Random Activation of Gene Expression', Nature Biotechnology, in press. Note that even though the cell type indicated is HT1080, since a random activation method was used, these sequence tags are not necessarily expressed in HT1080 under normal circumstances."
BASE COUNT 237 a 172 c 154 g 202 t
ORIGIN

alignment_scores:

Quality: 179.00 Length: 213
Ratio: 1.584 Gaps: 4
Percent Similarity: 53.052 Percent Identity: 25.822

alignment_block:

US-09-724-797-36 x BG182832 ..

Align seg 1/1 to: BG182832 from: 1 to: 765

```

190 GIuphealaasptYrValGluarGalaleuAlaArgArgArgGluG1 206
    ||| ..... ||| .....
60 GAAGTTACAAATTTTAAAGAAATCTGTAAAGATGAAAGAAAGCG 109
206 yglYglu.....AspleuAlaleuMetleuA 216
    ||| ..... ||| .....
110 CCTCGAAGATACACAAAGACCGAGTGGATTCTCTGAGCTGATGATG 159
216 spAlaHisasp.....ArgGlyLeuMetSerArgasn 226
    ||| ..... ||| .....
160 ACTCTCAAAATTCAAAAGAACTGAGTCCCAAAAGCTCTGTCCGATCTG 209
227 GluIleValSerThrValValThrPheIlePheThrGlyHisGluThrVa 243
    ||| ..... ||| .....
210 GAGCTGCTGCCCCAATCAATTATCTTTATTTGCTGCTATGAACCCAC 259
243 lAlaSerGlnValGlyAsnAlaValLeuSerLeuLeuAlaHisProaspG 260
    ||| ..... ||| .....
260 GAGCAGTGTCTCTCTTCATTATGTAAGAACTGGCCACTCACCCCTGATG 309
260 lInleuAspleuLeuArgArgProasp..... 269
    ||| ..... ||| .....
310 TCCAGCAAAACTGCAGAGGAATGTGATGCACTTTTACCAATAAGGCA 359
270 .....LeuAlaGlnAlaVala 275
    ||| ..... ||| .....
360 CCACCCACTATGATACTGCTGACAGATGAGATCTTGACATGAGTGGT 409
275 lGluGlyCysLeuArgTyrAspProSerValGlnSerAsnThrArgGlnL 292
    ||| ..... ||| .....
410 GAATGAACGCTCAGATTATTCCTCAATGTGATGAGACTTGAGAGGGTCT 459

```

292 euaspvalaspvalgluleuarglyargargleuargargaspval 308
 460 GCAAAAAAGATGTTGAGATCAATGGATGTTTCATCCCAAGGGCTGTG 509
 309 ValValValLeuAlaGlyAlaAlaAsnArgAspProArgArgTyrAsp 325
 510 GTATGATTCACAGCTATGCTCTTCACCGTGAACCAAGACTGACAG 559
 325 gproaspaspaspaspilleguarg..... 333
 560 GCCTGAGAAAGTTCCTCCCTGAAAGATTACGCAAGAACAGACAGACA 609
 334 . Aspprovalprosermetserphleglyalaglymetargtyrcysleu 349
 610 TAGATCTCTTCATATACACACCTTTGGAGTGGACCCAGAACTGCATT 659
 350 GlySerTyrLeuAlaArgThrGlnLeuArgAlaAlaVal 362
 660 GGCATGAGGTTTGCTCATGACATGAACCTGCTCTA 698

seq_name: gb_est1:AI323114

seq_documentation_block:

LOCUS AI323114 814 bp mRNA linear EST 23-DEC-1998
 DEFINITION mJ87b03.Y1 Soares mouse p3NMF19.5 Mus musculus cDNA clone
 IMAGE:483053 5, similar to gb:x63023 M.musculus mRNA for cytochrome
 P-45011A (MOUSE);, mRNA sequence.

ACCESSION AI323114
 VERSION AI323114.1 GI:4057543

KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus.

REFERENCE 1 (bases 1 to 814)
 AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
 Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
 Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
 Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
 Waterston,R.

TITLE The WashU-HMT Mouse EST Project
 JOURNAL Unpublished (1996)
 COMMENT Contact: Maria M/Mouse EST Project
 WashU-HMT Mouse EST Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: mouseest@wustl.edu

FEATURES
 source
 1. 814
 Location/Qualifiers
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /clone="IMAGE:483053"
 /clone_lib="Soares mouse p3NMF19.5"
 /dev_stage="19.5 dpc total fetus"
 /lab_host="DH10B (ampicillin resistant)"
 /note="Vector: pT73D (Pharmacia) with a modified
 polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
 was primed with a Not I - oligo(dT) primer [5'
 TGTATCCATTCGAGTGGAGCGGCGGATTTTTTTTTTTTTTTT 3']
 double-stranded cDNA was size selected, ligated to Eco RI
 adapters (Pharmacia), digested with Not I and cloned into
 the Not I and Eco RI sites of a modified pT73 vector
 (Pharmacia). Library went through one round of

normalization to a Cot = 5. Library constructed by Bento
 Soares and M.Fatima Bonaldo. RNA was kindly provided by
 Dr. Minou Ko (Wayne State University)."
 BASE COUNT 252 a 167 c 182 g 211 t 2 others
 ORIGIN

alignment_scores:
 Quality: 177.50 Length: 187
 Ratio: 1.723 Gaps: 5
 Percent Similarity: 55.080 Percent Identity: 28.342

alignment_block:

US-09-724-797-36 x AI323114 ..

Align seg 1/1 to: AI323114 from: 1 to: 814

209 AspleuLeuAlaLeuMetLeuAspAlaHisasp..... 219
 38 GATTTCTTCACAGCTGATGATTAACCTCCAGAAATTACAAACCAAGAAATC 87
 220 .ArgGlyLeuMetSerArgAsnGlnLeuValSerThrValValThrPheI 236
 88 TCATTAAGCATTAATCTGATGAGATTGTGGCTGACGACGATTCATT 137
 236 laphethrGlyHisGlnThrValAlaSerGlnValGlyAsnAlaValLeu 252
 138 TTTTGTCTGCTATGAGACACAAACAGTGCCTTCTTGCATTTGAT 187
 253 SerLeuLeuAlaHisPro.....AspGlnLeuAs 262
 188 TTGCTGGCATTCACCTGATGTACAGAAAGAACTTCAGATGAATGA 237
 262 P.....LeuLeuArgArgArgPro..... 268
 238 TCGCGCTGCCCCAATAGGCACTGCCACCTATATACCTGCTACGA 287
 269 .AspleuLeuAlaGlnAlaValGluGlyCysLeuArgTyrAspProSer 284
 288 TGGAGATATACACATGCGTGTGAATGAACCTCAGATTATATCCAAATC 337
 285 ValGlnSerAsnThrArgGlnLeuAspValAlaGlyLeuArgGlyAr 301
 338 GCTGGAAGACTTGAGAGGCTCTGTAAAGATGTTGAATCAATGGGCT 387
 301 gArgLeuArgArgAspAspValValValLeuAlaGlyAlaAlaAsnA 318
 388 ATTCATTTCCCAAGGAGCTGTGTGTATACCAACCTTGTCTTCAC 437
 318 rGAspProArgArgTyrAspArgProAspAspPheAspIlleguarg... 333
 438 AAGACCCGAAGTACTGGCCAGAGCCTGAGAGATTCGCCCTGAAGGATC 487
 334Aspprovalprosermetserphlegl 342
 488 AGCAAGAAAGATCANGACAGCATCAATCCTTACATGTACTCCCTTGG 537
 342 yAlaGlyMetArgTyrCysLeuGlySerTyrLeuAlaArgThrGlnLeu 359
 538 GATGGAACAAGAACTGCATTGGCATGAGGTTTGTCTCTCAATAAACATGA 587
 359 rGAlaAlaVal 362
 588 AAGTCTCTCTT 598
 seq_name: gb_est2:BG195942
 seq_documentation_block:
 LOCUS BG195942 828 bp mRNA linear EST 21-APR-2001
 DEFINITION R3115025 Athersys RAGE Library Homo sapiens cDNA, mRNA sequence.
 ACCESSION BG195942
 VERSION BG195942.1 GI:13717517
 KEYWORDS EST.


```

261  ....
261  GAGCAGGCTTCTCTCTTATTATGTATGAACCTGGCCACTCACCCTGATG 310
260  LIneuAspLeuArgArgArgProAsp..... 269
260  ...
311  TCCAGCAGAACTGCAGAGAAATGTGATGTTTACCCATATAGCA 360
270  .....LeuLeuAlaGlnAlaVal 275
361  CCACCCCTATGATATGCTGCTACAGATGAGTATCTTGACATGTGT 410
275  LgluGluCysLeuArgTyrAspProSerValGlnSerAsnThrArgGln 292
411  GAATGAAACGCTCAGATATTCCTCAATGCTATGAGACTTGAGAGGCTCT 460
292  euAspValAspValGluLeuArgGlyArgArgLeuArgArgAspAspVal 308
461  GCAAAAAAGATGTGTGAGATCAATGGGATGTTTCATCCCAAAGGGGTGTG 510
309  ValValValLeuAlaGlyAlaAlaAsnArgAspProArgArgTyrAspArg 325
511  GGGATGATTCACAGCTATGCTCTTCACCGTGACCCAAAGTACTGACAGA 560
325  gProAspAspPheAspIleGluArg..... 333
561  GCTGAGAGAGTCTCTCCCTGAAGAATTCACGACAGAGACAAAGACACA 610
334  .AspProValProSerMetSerPheGlyAlaGlyMetArgTyrCysLeu 349
611  TAGATCCTTACATATACACACCTTTGGAAATGGACCCAGAAACTGCATT 660
350  GlySerTyrLeuAlaArgThrGlnLeuArgAlaAlaVal 362
661  GGCATGAGGTTGTGCTCATGAAATGAACCTTGCTTA 699
```

